

Weighted information and entropy rates

Y. Suhov¹, I. Stuhl²

Abstract

The weighted entropy $H_\phi^w(X) = H_\phi^w(f)$ of a random variable X with values x and a probability-mass/density function f is defined as the mean value $\mathbb{E}I_\phi^w(X)$ of the weighted information $I_\phi^w(x) = -\phi(x) \log f(x)$. Here $x \mapsto \phi(x) \in \mathbb{R}$ is a given weight function (WF) indicating a 'value' of outcome x . For an n -component random vector $\mathbf{X}_0^{n-1} = (X_0, \dots, X_{n-1})$ produced by a random process $\mathbf{X} = (X_i, i \in \mathbb{Z})$, the weighted information $I_{\phi_n}^w(\mathbf{x}_0^{n-1})$ and weighted entropy $H_{\phi_n}^w(\mathbf{X}_0^{n-1})$ are defined similarly, with an WF $\phi_n(\mathbf{x}_0^{n-1})$. Two types of WFs ϕ_n are considered, based on additive and a multiplicative forms ($\phi_n(\mathbf{x}_0^{n-1}) = \sum_{i=0}^{n-1} \phi(x_i)$ and $\phi_n(\mathbf{x}_0^{n-1}) = \prod_{i=0}^{n-1} \phi(x_i)$, respectively). The focus is upon *rates* of the weighted entropy and information, regarded as parameters related to \mathbf{X} . We show that, in the context of ergodicity, a natural scale for an asymptotically additive/multiplicative WF is $\frac{1}{n^2} H_{\phi_n}^w(\mathbf{X}_0^{n-1})$ and $\frac{1}{n} \log H_{\phi_n}^w(\mathbf{X}_0^{n-1})$, respectively. This gives rise to *primary rates*. The next-order terms can also be identified, leading to *secondary rates*. We also consider emerging generalisations of the Shannon–McMillan–Breiman theorem.

1 Introduction

The purpose of this paper is to introduce and analyze weighted entropy rates for some basic random processes. In the case of a standard entropy, the entropy rate is a fundamental parameter leading to profound results and fruitful theories with far-reaching consequences, cf. [4]. The case of weighted entropies is much less developed, and this paper attempts to cover a number of aspects of this notion. In this work we treat two types of weight functions: additive and multiplicative. Conceptually, the present paper continues Refs [17, 18] and is connected with [15].

We work with a complete probability space $(\Omega, \mathfrak{B}, \mathbb{P})$ and consider random variables (RVs) as (measurable) functions $\Omega \rightarrow \mathcal{X}$ taking values in a measurable space $(\mathcal{X}, \mathfrak{M})$

2010 *Mathematics Subject Classification*: 60A10, 60B05, 60C05

Key words and phrases: weighted information and entropy, weight functions: additive and multiplicative, rates, ergodic processes, Markov chains, Gaussian processes, topological entropy and pressure

¹ Mathematics Dept., Penn State University, University Park, State College, PA 16802, USA; DPMMS, University of Cambridge, UK;

E-mail: yms@statslab.cam.ac.uk

² Mathematics Dept., University of Denver, Denver, CO 80208 USA; Appl. Math and Prob. Theory Dept., University of Debrecen, Debrecen, 4028, HUN;

E-mail: izabella.stuhl@du.edu

equipped with a countably additive reference measure ν . Probability mass functions (PMFs) or probability density functions (PDFs) are defined relative to ν . (The difference between PMFs (discrete parts of probability measures) and PDFs (continuous parts) is insignificant for most of the work; this will be reflected in a common acronym PM/DF.) In the case of an RV collection $\{X_i\}$, the space of values \mathcal{X}_i and the reference measure ν_i may vary with i . (Some of the X_i may be random vectors.)

Given a (measurable) function $x \in \mathcal{X} \mapsto \phi(x) \in \mathbb{R}$, and an RV $X : \Omega \rightarrow \mathcal{X}$, with a PM/DF f , the weighted information (WI) $I_\phi^w(x)$ with weight function (WF) ϕ contained in an outcome $x \in \mathcal{X}$ is given by

$$I_\phi^w(x) = -\phi(x) \log f(x). \quad (1.1)$$

The symbol $I_\phi^w(X)$ is used for the random WI, under PM/DF f . Next, one defines the weighted entropy (WE) $h_\phi^w(f)$ of f (or X) as

$$h_\phi^w(f) = - \int_{\mathcal{X}} \phi(x) f(x) \log f(x) \nu(dx) = \mathbb{E} I_\phi^w(X) \quad (1.2)$$

whenever the integral $\int_{\mathcal{X}} |\phi(x)| f(x) |\log f(x)| \nu(dx) < \infty$. (A common agreement $0 = 0 \cdot \log 0 = 0 \cdot \log \infty$ is in place throughout the paper.) Here and below we denote by \mathbb{E} the expectation relative to \mathbb{P} (or an induced probability measure emerging in a given context). For $\phi(x) \geq 0$, the WE in a discrete case (when \mathcal{X} is a finite or a countable set) is non-negative. For $\phi(x) = 1$, we obtain the standard information $I(x) = -\log f(x)$ (SI) and standard entropy $h(f) = \mathbb{E} I(X)$ (SE).

Let $\mathbf{X}_0^{n-1} = (X_0, X_1, \dots, X_{n-1})$ be a random vector (string), with components $X_i : \Omega \rightarrow \mathcal{X}_i$, $0 \leq i \leq n-1$. Let $f_n(\mathbf{x}_0^{n-1})$ be the joint PM/DF relative to measure $\nu_0^{n-1}(d\mathbf{x}_0^{n-1}) = \prod_{i=0}^{n-1} \nu_i(dx_i)$ where $\mathbf{x}_0^{n-1} = (x_0, \dots, x_{n-1}) \in \prod_{i=0}^{n-1} \mathcal{X}_i := \mathcal{X}_0^{n-1}$. Given a function $\mathbf{x}_0^{n-1} \in \mathcal{X}_0^{n-1} \mapsto \phi_n(\mathbf{x}_0^{n-1}) \in \mathbb{R}$, the joint WE $h_{\phi_n}^w(f_n)$ of X_0, \dots, X_{n-1} with WF ϕ_n is given by

$$h_{\phi_n}^w(f_n) = - \int_{\mathcal{X}_0^{n-1}} \phi_n(\mathbf{x}_0^{n-1}) f_n(\mathbf{x}_0^{n-1}) \log f_n(\mathbf{x}_0^{n-1}) \nu_0^{n-1}(d\mathbf{x}_0^{n-1}) = \mathbb{E} I_{\phi_n}^w(\mathbf{X}_0^{n-1}), \quad (1.3)$$

where $I_{\phi_n}^w(\mathbf{x}_0^{n-1})$ represents the WI in the joint outcome \mathbf{x}_0^{n-1} :

$$I_{\phi_n}^w(\mathbf{x}_0^{n-1}) = -\phi_n(\mathbf{x}_0^{n-1}) \log f_n(\mathbf{x}_0^{n-1}). \quad (1.4)$$

We focus upon two kinds of weight functions $\phi_n(\mathbf{X}_0^{n-1})$: additive and multiplicative, and their asymptotical modifications. Both relate to the situation where $\mathbf{X}_0^{n-1} = (X_0, \dots, X_{n-1})$ and each component X_j takes values in the same space: $\mathcal{X}_j = \mathcal{X}_1 = \mathcal{X}$. In the simplest form, additivity and multiplicativity mean representations

$$\phi_n(\mathbf{x}_0^{n-1}) = \sum_{0 \leq j < n} \varphi(x_j) \quad \text{and} \quad \phi_n(\mathbf{x}_0^{n-1}) = \prod_{0 \leq j < n} \varphi(x_j), \quad (1.5)$$

where $x \in \mathcal{X} \mapsto \varphi(x)$ is a given functions (one-digit WFs). In the additive case we can allow φ to be of both signs whereas in the multiplicative case we suppose $\varphi \geq 0$.

Additive weight functions may emerge in relatively stable situations where each observed digit X_j brings reward or loss $\varphi(X_j)$ (bearing opposite signs); the summatory value $\phi_n(\mathbf{X}_0^{n-1})$ is treated as a cumulative gain or deficit after n trials. Multiplicative weight functions reflect a more turbulent scenario where the value $\phi_n(\mathbf{X}_0^{n-1})$ increases/decreases by a factor $\varphi(X_n)$ when outcome X_n is observed. Cf. [16]. As before, for $\phi_n(\mathbf{x}_0^{n-1}) \equiv 1$ we obtain the SE $h(f_n)$ and SI $I(\mathbf{x}_0^{n-1})$.

Our goal is to introduce concepts of *rates* for $h_{\phi_n}^w(f_n)$ and $I_{\phi_n}^w(\mathbf{x}_0^{n-1})$ characterising the order of growth/decay as $n \rightarrow \infty$. To this end we consider a (discrete-time) random process $\mathbf{X} = (X_i, i \in \mathbb{Z})$ or $\mathbf{X} = (X_i, i \in \mathbb{Z}_+)$, with a probability distribution \mathbb{P} ; vector \mathbf{X}_0^{n-1} will represent an initial string generated by the process. In the case of the SE and SI, the rates are defined as $\lim_{n \rightarrow \infty} \frac{1}{n} h(f_n)$ and $\lim_{n \rightarrow \infty} \frac{1}{n} I(\mathbf{x}_0^{n-1})$, and for an ergodic process they coincide almost everywhere relative to the distribution \mathbb{P} . See [1], [3], [4]. For the WE and WI we find it natural to introduce *primary* and *secondary* rates. The former emerges as a limit of $\frac{1}{n^2} H_{\phi_n}^w(\mathbf{X}_0^{n-1})$ for asymptotically additive WFs and of $\frac{1}{n} \log H_{\phi_n}^w(\mathbf{X}_0^{n-1})$ for asymptotically multiplicative WFs. The secondary rate, roughly, provides a ‘correction term’, although in a number of situations (when the primary rate vanishes) the secondary rate should bear a deeper significance. We also consider generalisations of the Shannon–McMillan–Breiman (SMB) theorem for asymptotically additive WFs.

The paper is organised as follows. In Section 2 we put forward the concepts of asymptotically additive and multiplicative WFs. In Section 3, the primary and secondary rates for additive case are discussed. Section 3 ...

2 Asymptotic additivity and multiplicativity

Here we introduce classes of asymptotically additive and multiplicative WFs for which we develop results on rates in the subsequent sections. The object of study is a discrete-time random process $\mathbf{X}_0^\infty = (X_n : n \in \mathbb{Z}_+)$ or $\mathbf{X} = (X_n : n \in \mathbb{Z})$. We begin with a simple example where \mathbf{X}_0^∞ is an IID (Bernoulli) process with values in \mathcal{X} : here, for $\mathbf{x}_0^{n-1} = (x_0, \dots, x_{n-1}) \in \mathcal{X}^n$, the joint PM/DF for string $\mathbf{X}_0^{n-1} = (X_0, \dots, X_{n-1})$ is $f_n(\mathbf{x}_0^{n-1}) = \prod_{i=0}^{n-1} p(x_i)$ where $p(x) = p_0(x)$ is the one-time marginal PM/DF, $x \in \mathcal{X}$. We start with a straightforward remark:

- (a) For a sequence of IID random variables \mathbf{X}_0^∞ and an additive WF $\phi_n(\mathbf{X}_0^{n-1}) = \sum_{0 \leq j < n} \varphi(X_j)$, the WI has a representation:

$$I_{\phi_n}^w(\mathbf{X}_0^{n-1}) = -\phi(\mathbf{X}_0^{n-1}) \log f_n(\mathbf{X}_0^{n-1}) = -\sum_{j=0}^{n-1} \varphi(X_j) \sum_{l=0}^{n-1} \log p(X_l). \quad (2.1)$$

Next, with $H(p) = -\mathbb{E}[\log p(X)]$ and $H_\varphi^w(p) = -\mathbb{E}[\varphi(X) \log p(X)]$ (the one-digit SE and WE, respectively):

$$H_{\phi_n}^w(f_n) = n(n-1)H(p)\mathbb{E}[\varphi(X)] + nH_\varphi^w(p) := n(n-1)A_0 + nA_1. \quad (2.2)$$

- (b) For a sequence of IID random variables \mathbf{X}_0^∞ and a multiplicative WF $\phi_n(\mathbf{X}_0^{n-1})$
 $= \prod_{0 \leq j < n} \varphi(X_j)$:

$$I_{\phi_n}^w(\mathbf{X}_0^{n-1}) = -\phi_n(\mathbf{X}_0^{n-1}) \log f_n(\mathbf{X}_0^{n-1}) = -\prod_{j=0}^{n-1} \varphi(X_j) \sum_{l=0}^{n-1} \log p(X_l). \quad (2.3)$$

Next,

$$H_{\phi_n}^w(f_n) = nH_\varphi^w(p) [\mathbb{E}\varphi(X)]^{n-1} := B_0^{n-1} \times nB_1. \quad (2.4)$$

Values A_0 and B_0 are referred to as primary rates and A_1 and B_1 as secondary rates.

Eqns (2.1)–(2.4) provide intuition for formulas of convergence (2.14)–(2.15) which yield versions of the SMB theorem for the WI and WE in a general case with asymptotically additive WFs. (A number of subsequent results will be established or illustrated under specific restrictions, viz., Markovian or Gaussian assumptions.) We consider $\mathcal{X} = \mathcal{X}^\mathbb{Z}$ (the space of trajectories over \mathbb{Z}) and $\mathcal{X}_+ = \mathcal{X}^{\mathbb{Z}_+}$ (the set of trajectories over \mathbb{Z}_+), equipped with the corresponding sigma-algebras. As was said, symbol \mathbb{P} is used for a probability measure on \mathcal{X}_+ or \mathcal{X} generated by process \mathbf{X}_0^∞ or \mathbf{X} . (In the case of \mathbf{X} , symbol \mathbb{P} will be related to a stationary process, while for \mathbf{X}_0^∞ some alternative possibilities can be considered as well, involving initial conditions.) Symbol \mathbb{E} refers to the expectation relative to \mathbb{P} . Next, L_2 stands for the Hilbert space $L_2(\mathcal{X}_+, \mathbb{P})$ or $L_2(\mathcal{X}, \mathbb{P})$ and L_1 for the space $L_1(\mathcal{X}_+, \mathbb{P})$ or $L_1(\mathcal{X}, \mathbb{P})$. The joint PM/DF for string \mathbf{X}_0^{n-1} is again denoted by f_n : $f_n(\mathbf{x}_0^{n-1}) = \frac{\mathbb{P}(\mathbf{X}_0^{n-1} \in d\mathbf{x}_0^{n-1})}{\nu^n(d\mathbf{x}_0^{n-1})}$. The focus will be upon rates of the WI $I_{\phi_n}^w(\mathbf{X}_0^{n-1})$ and WE $H_{\phi_n}^w(f_n)$; see (1.4) and (1.3).

One of aspects of this work is to outline general classes of WFs ϕ_n and RPs \mathbf{X} , replacing the exact formulas in (2.2) and (2.4) by suitable asymptotic representations (with emerging asymptotic counterparts of parameters A_0 , A_1 and B_0 , B_1). In our opinion, a natural class of RPs here are ergodic processes; a part of the assertions in this paper are established in this class. The basis for such a view is that for an ergodic RP $\mathbf{X} = (X_n, n \in \mathbb{Z})$ the limit

$$\lim_{n \rightarrow \infty} \frac{-1}{n} \log f_n(\mathbf{X}_0^{n-1}) = h \quad (2.5)$$

exists \mathbb{P} -a.s. according to results by Barron (1985) [3] and Algoet–Cover (1988) [1]. Cf., e.g., [1], Theorem 2, and the biblio therein. The limiting value h is identified as the SE rate of RP \mathbf{X} (the SMB theorem). However, a number of properties in the present paper are proven under Markovian assumptions, due to technical complications. In some situations (for Gaussian processes) we are able to analyse the situation without referring directly to ergodicity (or stationarity).

Another aspect is related to suitable assumptions upon WFs. One assumption is that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \phi_n(\mathbf{X}_0^{n-1}) = \alpha, \quad \mathbb{P}\text{-a.s. and/or in } L_2 \quad (\text{asymptotic additivity}); \quad (2.6)$$

together with (2.5) it leads to identification of the primary rate A_0 :

$$A_0 = \alpha h. \quad (2.7)$$

The impact of process \mathbf{X} in assumption (2.6) is reduced to the form of convergence (\mathbb{P} -a.s. or $L_2(\mathcal{X}, \mathbb{P})$). A stronger tie between ϕ_n and \mathbf{X} is introduced in an asymptotic relation (2.8) arising from (2.2):

$$\lim_{n \rightarrow \infty} \frac{1}{n} H_{\phi_n}^w(f_n) = A_1. \quad (2.8)$$

An instructive property implying (2.8) is that $\forall j \in \mathbb{Z}$,

$$\lim_{n \rightarrow \infty} \mathbb{E}[\phi_n(\mathbf{X}_0^{n-1}) \log p^{(j)}(X_j | \mathbf{X}_0^{j-1})] = A_1 \quad (\text{strong asymptotic additivity}). \quad (2.9)$$

This yields an identification of the secondary rate A_1 . Here and below, $p^{(j)}(y | \mathbf{x}_0^{j-1})$ represents the conditional PM/DF of having $X_j = y$ given that string \mathbf{X}_0^{j-1} coincides with \mathbf{x}_0^{j-1} ; see Eqn (3.2) below. Assumptions (2.6) and (2.9) are relevant in Section 3, Theorem 3.1.

An informal meaning of (2.6) is that there is an approximation

$$\frac{\phi_n(\mathbf{x}_0^{n-1}) - \phi_n^*(\mathbf{x})}{n} \rightarrow 0 \quad \text{where} \quad \phi_n^*(\mathbf{x}) = \sum_{j=0}^{n-1} \varphi^*(S^j \mathbf{x}), \quad (2.10)$$

for some measurable function $\mathbf{x} \in \mathcal{X} \mapsto \varphi^*(\mathbf{x}) \in \mathbb{R}$ from L_1 , with $\alpha = \mathbb{E}\varphi^*(\mathbf{X})$. Here and below, S stands for the shift in \mathcal{X} : $(S^j \mathbf{x})_l = x_{l-j}$ for $\mathbf{x} = (x_l) \in \mathcal{X}$. From this point of view, condition (2.8) is instructive when $A_0 = 0$ (i.e., h or α vanishes).

Let us now pass to multiplicative WFs. An assumption used in Section 4, Theorem 5, claims that

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{1}{n} \log \phi_n(\mathbf{X}_0^{n-1}) &= \log \beta, \quad \text{or, equivalently,} \\ \lim_{n \rightarrow \infty} [\phi_n(\mathbf{X}_0^{n-1})]^{1/n} &= \beta, \quad \mathbb{P}\text{-a.s. (asymptotic multiplicativity).} \end{aligned} \quad (2.11)$$

Similarly to (2.10), Eqn (2.11) means, essentially, that

$$\left[\frac{\phi_n(\mathbf{X}_0^{n-1})}{\phi_n^*(\mathbf{X})} \right]^{1/n} \rightarrow 1 \quad \text{where} \quad \phi_n^*(\mathbf{x}) = \prod_{0 \leq j < n} \varphi^*(S^j \mathbf{x}), \quad (2.12)$$

for some measurable function $\mathbf{x} \in \mathcal{X} \mapsto \varphi^*(\mathbf{x}) > 0$, with $(\log \varphi^*) \in L_1$ and $\mathbb{E} \log \varphi^*(\mathbf{X}) = \beta$. A stronger form of such a condition is an exact equality: $\phi_n(\mathbf{x}_0^{n-1}) = \prod_{0 \leq j < n} \varphi(x_j)$; cf. (1.5).

For a future use, we suggest an integral form of condition (2.12): as $n \rightarrow \infty$,

$$\left\{ \frac{\mathbb{E}[\phi_n(\mathbf{X}_0^{n-1}) \log f_n(\mathbf{X}_0^{n-1})]}{\mathbb{E}[\phi_n^*(\mathbf{X}) \log f_n(\mathbf{X}_0^{n-1})]} \right\}^{1/n} \rightarrow 1, \quad \text{or} \quad \frac{1}{n} \log \frac{\mathbb{E}[\phi_n(\mathbf{X}_0^{n-1}) \log f_n(\mathbf{X}_0^{n-1})]}{\mathbb{E}[\phi_n^*(\mathbf{X}) \log f_n(\mathbf{X}_0^{n-1})]} \rightarrow 0. \quad (2.13)$$

The main results of this paper can be described as follows.

(A) For additive or asymptotically additive WFs (i.e., under assumption (1.5) or (2.6)) we analyse the limits

$$(i) \quad A_0 = \lim_{n \rightarrow \infty} \frac{I_{\phi_n}^w(\mathbf{X}_0^{n-1})}{n^2}, \quad (ii) \quad A_0 = \lim_{n \rightarrow \infty} \frac{H_{\phi_n}^w(f_n)}{n^2}. \quad (2.14)$$

(B) For multiplicative or asymptotically multiplicative WFs (i.e., under assumptions (1.5) or (2.11)), the focus will be on convergences

$$(i) \quad \bar{B}_0 = \lim_{n \rightarrow \infty} \frac{1}{n} \log I_{\phi_n}^w(\mathbf{X}_0^{n-1}), \quad (ii) \quad B_0 = \lim_{n \rightarrow \infty} \frac{1}{n} \log H_{\phi_n}^w(f_n). \quad (2.15)$$

In (2.14i), (2.15i) we bear in mind various forms of convergence for random variables (see specific statements below). For multiplicative WFs we will also identify an analog of the value B_1 from (2.4) for Markov chains:

$$B_1 = \lim_{n \rightarrow \infty} \frac{H_{\phi_n}^w(f_n)}{n B_0^{n-1}}. \quad (2.16)$$

We want to stress that some properties are established in this paper under rather restrictive assumptions, although in our opinion, a natural class of RPs for which these properties hold is much wider. This view is partially supported by an analysis of Gaussian processes \mathbf{X}_0^∞ is conducted in Section 5.

Remark 2.1 The normalisation considered in (2.8), (2.14) and (2.15) is connected with stationarity/ergodicity of RP \mathbf{X} and various forms of asymptotic additivity and multiplicativity of WFs ϕ_n . Abandoning these types of assumptions may lead to different types of scaling.

3 Rates for additive WFs

3.1 A general statement

Consider first a general case where \mathbf{X} is a stationary ergodic RP with a probability distribution \mathbb{P} on \mathcal{X} . In this case we write

$$I_{\phi_n}^w(\mathbf{x}_0^{n-1}) = -\phi_n(\mathbf{x}_0^{n-1}) \left[\log p_0(x_0) + \sum_{1 \leq j < n} \log p^{(j)}(x_j | \mathbf{x}_0^{j-1}) \right]. \quad (3.1)$$

As in Eqn (2.9), $p^{(j)}(y | \mathbf{x}_0^{j-1})$ represents the conditional PM/DF of having $X_j = y$ given that string \mathbf{X}_0^{j-1} coincides with \mathbf{x}_0^{j-1} , and $p_0(y)$ is the PM/DF for X_0 :

$$p_0(y) = \frac{\mathbb{P}(X_0 \in dy)}{\nu(dy)}, \quad p^{(j)}(y | \mathbf{x}_0^{j-1}) = \frac{\mathbb{P}(X_j \in dx)}{\nu(dx)}, \quad y \in \mathcal{X}, \quad \mathbf{x}_0^{j-1} \in \mathcal{X}^j. \quad (3.2)$$

The SE rate h is defined by

$$h = -\mathbb{E} \log p(X_0 | \mathbf{X}_{-\infty}^{-1}) \quad (3.3)$$

where $p(y | \mathbf{x}_{-\infty}^{-1})$ is the conditional PM/DF for $X_0 = y$ given $\mathbf{x}_{-\infty}^{-1}$, an infinite past realization of \mathbf{X} . As before, set $H_{\phi_n}^w(f_n) = \mathbb{E} I_{\phi_n}^w(\mathbf{X}_0^{n-1})$. Recall, the SMB theorem asserts that for an ergodic RP \mathbf{X} , the following limit exists \mathbb{P} -a.s.

$$\lim_{n \rightarrow \infty} \frac{1}{n} \left[\log p_0(X_0) + \sum_{1 \leq j < n} \log p^{(j)}(X_j | \mathbf{X}_0^{j-1}) \right] = h. \quad (3.4)$$

Theorem 3.1 *Given an ergodic probability distribution \mathbb{P} on \mathcal{X} , consider the WI $I_{\phi_n}^w(\mathbf{X}_0^{n-1})$ and the WE $H_{\phi_n}^w(f_n)$ as defined in (1.4) and (1.3). Suppose that convergence in (2.6) holds \mathbb{P} -a.s. Then:*

- (I) *Convergence in (2.14 i) holds true, \mathbb{P} -a.s., with $A_0 = \alpha h$ where α is as in (2.6) and h as in (3.3). That is:*

$$\lim_{n \rightarrow \infty} \frac{I_{\phi_n}^w(\mathbf{X}_0^{n-1})}{n^2} = \alpha h, \quad \mathbb{P}\text{-a.s.}$$

- (II) *Furthermore, (a) suppose that the WFs ϕ_n exhibit convergence (2.6), \mathbb{P} -a.s., with a finite α , and $|\phi_n(\mathbf{X}_0^{n-1})/n| \leq c$ where c is a constant independent of n . Suppose also that convergence in Eqn (2.5) holds with $h \in [0, \infty)$ given by (3.3). Then convergence in (2.14 ii) holds true, as before with $A_0 = \alpha h$:*

$$\lim_{n \rightarrow \infty} \frac{H_{\phi_n}^w(f_n)}{n^2} = \alpha h.$$

- (b) *Likewise, convergence in Eqn (2.14 ii) holds true whenever convergences (2.6) and (2.5) hold \mathbb{P} -a.s. and $|\log f_n(\mathbf{X}_0^{n-1})/n| \leq c$ where c is a constant. Finally, (c) suppose that convergence in (2.6) and (2.5) holds in L_2 , with finite α and h . Then convergence in (2.14 ii) holds true, again with $A_0 = \alpha h$.*

Proof. Assertion (I) follows immediately from the \mathbb{P} -a.s. convergence in Eqns (2.6) and (3.3). The same is true of assertions (IIa) and (IIb), with the help of the Lebesgue dominated convergence theorem. Assertion (IIc) follows from the L_2 -convergence and continuity of the scalar product. \blacksquare

Remark 3.2 The assumption in statement (IIc) of Theorem 3.1 that the limit in (2.5) holds in L_2 (i.e., an L_2 -SMB theorem) can be checked in a number of special cases. We conjecture that a sufficient condition is that \mathbb{P} is ergodic and RV $\log p(X_0|\mathbf{X}_{-\infty}^{-1})$ lies in L_2 . However, to the best of our knowledge, it is an open question. The fact that the limits in parts (I) and (IIa) coincide can be considered as an analog of the SMB theorem to the case under consideration.

Remark 3.3 Under conditions of Theorem 3.1, the bound $|\log f_n(\mathbf{X}_0^{n-1})/n| \leq c$ in assertion (b) holds when \mathcal{X} is a finite or a countable set (the Chung-Neveu lemma).

Remark 3.4 The factor $\frac{1}{n}$ in assumption (2.6) can be replaced by $\frac{1}{a(n)}$ where $a(n)$ is a given increasing sequence of positive numbers. In this case we can speak of a moderated asymptotic additivity of WF ϕ_n . Accordingly, in (2.14) the denominator n^2 should be replaced with $na(n)$.

Remark 3.5 The statement of Theorem 3.1 remains in force when in representation (3.1) the sum $\log p_0(X_0) + \sum_{1 \leq j < n} \log p^{(j)}(X_j|\mathbf{X}_0^{j-1})$ is replaced with $\sum_{j=0}^{n-1} \log p(X_j|\mathbf{X}_{-\infty}^{j-1})$ and/or WF $\phi_n(\mathbf{X}_0^{n-1})$ is replaced by the sum $\phi_n^*(\mathbf{X}) = \sum_{j=0}^{n-1} \varphi^*(S^j \mathbf{X})$ (cf. Eqn (2.10)), under appropriate assumptions upon φ^* . This is achieved by making use of standard Ergodic theorems (Birkhoff and von Neumann).

3.2 The Markovian case

It is instructive to affiliate an assertion analogous to Theorem 3.1 for a Markov chain of order $k \geq 1$. In this case the PM/DF $f_n(\mathbf{x}_0^{n-1})$, relative to reference measure ν^k on \mathcal{X}^k , for $n > k$ has the form

$$f_n(\mathbf{x}_0^{n-1}) = \lambda(\mathbf{x}_0^{k-1}) \prod_{0 \leq j < n-k} p(x_{j+k} | \mathbf{x}_j^{j+k-1}). \quad (3.5)$$

Here λ yields a PM/DF for an initial string: $\lambda(\mathbf{x}_0^{k-1}) \geq 0$ and $\int_{\mathcal{X}^k} \lambda(\mathbf{x}_0^{k-1}) \nu^k(d\mathbf{x}_0^{k-1}) = 1$.

Further, as above, $p(y | \mathbf{x}_j^{j+k-1})$ represents the conditional PM/DF of having $X_{j+k} = y$ given that string \mathbf{X}_j^{j+k-1} coincides with \mathbf{x}_j^{j+k-1} . Next, let π be an equilibrium PM/DF on \mathcal{X}^k , with

$$\pi(\mathbf{x}_0^{k-1}) = \int_{\mathcal{X}} \pi(x' \vee \mathbf{x}_0^{k-2}) p(x_{k-1} | x' \vee \mathbf{x}_0^{k-2}) \nu(dx') \quad (3.6)$$

where string $x' \vee \mathbf{x}_0^{k-1} = (x', x_0, \dots, x_{k-2}) \in \mathcal{X}^k$. Denote by \mathbb{P}_λ and $\mathbb{P} = \mathbb{P}_\pi$ the probability distributions (on \mathcal{X}_+ and \mathcal{X} , respectively) generated by the process with initial PM/DF λ and π . Further, let \mathbb{E} and \mathbb{E}_λ stand for the expectations under \mathbb{P} and \mathbb{P}_λ . Set

$$h = -\mathbb{E} \log p(X_k | \mathbf{X}_0^{k-1}) = - \int_{\mathcal{X}^k} \pi(\mathbf{x}_0^{k-1}) p^{(k)}(x_k | \mathbf{x}_0^{k-1}) \log p^{(k)}(x_k | \mathbf{x}_0^{k-1}) \nu^k(d\mathbf{x}_0^{k-1}). \quad (3.7)$$

Next, define $H_{\phi_n}^w(f_n, \pi) = \mathbb{E} I_{\phi_n}^w(\mathbf{X}_0^{n-1})$ and $H_{\phi_n}^w(f_n, \lambda) = \mathbb{E}_\lambda I_{\phi_n}^w(\mathbf{X}_0^{n-1}, \lambda)$ where

$$I_{\phi_n}^w(\mathbf{x}_0^{n-1}, \lambda) = -\phi_n(\mathbf{x}_0^{n-1}) \left[\log \lambda(\mathbf{x}_0^{k-1}) + \sum_{0 \leq j < n-k} \log p^{(j+k)}(x_{j+k} | \mathbf{x}_j^{j+k-1}) \right]. \quad (3.8)$$

For definiteness, in Theorem 3.2 below we adopt conditions in a rather strong form, without distinguishing between different possibilities listed in the body of Theorem 3.1. The proof of Theorem 3.2 is essentially a repetition of that of Theorem 3.1, with an additional help from the Ergodic theorems.

Theorem 3.6 *Let \mathbf{X}_0^∞ be a k -order Markov chain with an initial PM/DF $\lambda(\mathbf{x}_0^{k-1})$ where $k \geq 1$. Assume that (i) Eqn (2.6) is fulfilled, both in L_2 and \mathbb{P} -a.s., (ii) the stationary probability measure \mathbb{P} on \mathcal{X} is ergodic, (iii) $\log \lambda(\mathbf{X}_0^{k-1})$ and $\log p(X_k | \mathbf{X}_0^{k-1})$ belong to L_2 , (iv) $\text{supp } \lambda \subseteq \text{supp } \pi$. Then the limiting relations (2.14) are satisfied, for both choices of $I_{\phi_n}^w(\mathbf{x}_0^{n-1})$, $H_{\phi_n}^w(f_n)$ and of $I_{\phi_n}^w(\mathbf{x}_0^{n-1}, \lambda)$, $H_{\phi_n}^w(f_n, \lambda)$, with $A = \alpha h$ where α is as in (2.6) and h as in (3.7). Correspondingly, convergence in (2.14 i) holds \mathbb{P} -a.s. and \mathbb{P}_λ -a.s.*

A similar assertion could be given in the case of a general initial probability distribution $\lambda(d\mathbf{x}_0^{k-1})$ on \mathcal{X}^k which can be singular relative to ν^k . Here, for $n > k$ we consider the PM/DF $f_n(\mathbf{x}_0^{n-1})$ with respect to $\lambda(d\mathbf{x}_0^{k-1}) \nu^{n-k}(d\mathbf{x}_k^{n-1})$ on \mathcal{X}^n :

$$f_n(\mathbf{x}_0^{n-1}) = \prod_{1 \leq j < n} p(x_j | \mathbf{x}_{j-k}^{j-1}), \quad \text{with} \quad \int_{\mathcal{X}^n} f_n(\mathbf{x}_0^{n-1}) \lambda(d\mathbf{x}_0^{k-1}) \nu^{n-k}(d\mathbf{x}_k^{n-1}) = 1. \quad (3.9)$$

Then \mathbb{P}_λ denotes the probability distribution (on \mathcal{X}_+) generated by the process with the initial distribution λ whereas \mathbb{E}_λ stands for the expectation under \mathbb{P}_λ . The notation

$\mathbb{P} = \mathbb{P}_\pi$ and $\mathbb{E} = \mathbb{E}_\pi$ has the same meaning as before, with $\pi(\mathbf{x}_0^{k-1})$ being an equilibrium PM/DF relative to ν^k on \mathcal{X}^k . Accordingly, we now define

$$I_{\phi_n}^w(\mathbf{x}_0^{n-1}) = -\phi_n(\mathbf{x}_0^{n-1}) \sum_{0 \leq j < n-k} \log p^{(j+k)}(x_{j+k} | \mathbf{x}_j^{j+k-1}) \quad (3.10)$$

and $H_{\phi_n}^w(f_n, \boldsymbol{\lambda}) = \mathbb{E}_{\boldsymbol{\lambda}} I_{\phi_n}^w(\mathbf{X}_0^{n-1})$.

Theorem 3.7 *Let \mathbf{X}_0^∞ be a k -order Markov chain with an initial probability measure $\boldsymbol{\lambda}(\mathrm{d}\mathbf{x}_0^{k-1})$ where $k \geq 1$. Adopt assumptions (i) and (ii) of Theorem 3.2. In addition, suppose that (iii) $p^{(k)}(X_k | \mathbf{X}_0^{k-1}) > 0$ \mathbb{P} -a.s. (implying that $\pi(\mathbf{x}_0^{k-1})$ is strictly positive ν^k -a.s. on \mathcal{X}^k) and that $\log p^{(k)}(X_k | \mathbf{X}_0^{k-1})$ belongs to L_2 . With $I_{\phi_n}^w(\mathbf{x}_0^{n-1})$ as in (3.10), the assertions of Theorem 3.2 hold true, mutatis mutandis, and convergence in (2.14 i) takes place \mathbb{P} -a.s. and $\mathbb{P}_{\boldsymbol{\lambda}}$ -a.s. Furthermore, convergence in (2.14 ii) holds for both $H_{\phi_n}^w(f_n)$ and $H_{\phi_n}^w(f_n, \boldsymbol{\lambda})$.*

Theorem 3.8 *Suppose that $\phi_n(\mathbf{x}_0^{n-1}) = \sum_{j=0}^{n-1} \varphi(x_j)$. Let \mathbf{X} be a stationary RP with the property that $\forall i \in \mathbb{Z}$ there exists the limit*

$$\begin{aligned} -A_1 &:= \lim_{n \rightarrow \infty} \sum_{j \in \mathbb{Z}: |j+i| \leq n} \mathbb{E}[\varphi(X_0) \log p^{(n+i+j)}(X_j | \mathbf{X}_{-n-i}^{j-1})] \\ &= \sum_{j \in \mathbb{Z}} \mathbb{E}[\varphi(X_0) \log p(X_j | \mathbf{X}_{-\infty}^{j-1})], \end{aligned} \quad (3.11)$$

and the last series converges absolutely. Then $\lim_{n \rightarrow \infty} \frac{1}{n} H_{\phi_n}^w(f_n) = A_1$.

Proof. Set: $n_1 = -[n/2]$, $n_2 = [(n+1)/2] - 1$. Then we can write

$$\begin{aligned} \frac{1}{n} H_{\phi_n}^w(f_n) &= -\frac{1}{n} \sum_{n_1 \leq i \leq n_2} E_{n,i} \\ \text{where } E_{n,i} &= \sum_{l=n_1-i}^{n_2-i} \mathbb{E}[\varphi(X_0) \log p^{(-n_1+l+i)}(X_l | \mathbf{X}_{n_1-i}^{l-1})]. \end{aligned} \quad (3.12)$$

(For $l = n_1 - i$, we have the term $\log p_0(X_{n_1-i})$.) By virtue of (3.11), each $E_{n,i}$ tends to $-A_1$, hence the Cesaro mean does too. \blacksquare

Remark 3.9 Condition (3.11) alludes that $\mathbb{E}\varphi(X_i) = 0$. We will now show that (3.11) holds when \mathcal{X} is a finite set and \mathbf{X} is a stationary ergodic Markov chain with positive transition probabilities $\mathbf{p}(x, y)$ and equilibrium probabilities $\pi(x)$, $x, y \in \mathcal{X}$. Then $\rho := \min \mathbf{p}(x, y)$ satisfies $0 < \rho < 1$, and the s -step transition probabilities $\mathbf{p}^{(s)}(x, y)$ obey $|\mathbf{p}^{(s)}(x, y) - \pi(y)| \leq 2(1 - \rho)^s$ (a Doeblin property). Assume that $\mathbb{E}\varphi(X_i) = \sum_{x \in \mathcal{X}} \pi(x) \varphi(x) = 0$. Then, $\forall n_1, n_2 \in \mathbb{Z}$ with $n_1 < 0 < n_2$,

$$\begin{aligned} &\sum_{j \in \mathbb{Z}: n_1 \leq j \leq n_2} \mathbb{E}[\varphi(X_0) \log p_{j-n_1}(X_j | \mathbf{X}_{n_1}^{j-1})] \\ &= \sum_{n_1 \leq j \leq 0} \sum_{x, y, z \in \mathcal{X}} [\pi(x) \mathbf{p}(x, y) \log \mathbf{p}(x, y) \mathbf{p}^{(-j)}(y, z) \varphi(z)] \\ &\quad + \sum_{0 \leq j < n_2} \sum_{x, y, z \in \mathcal{X}} [\varphi(x) \pi(x) \mathbf{p}^{(j-1)}(x, y) \mathbf{p}(y, z) \log \mathbf{p}(y, z)]. \end{aligned} \quad (3.13)$$

As $-n_1, n_2 \rightarrow \infty$, the RHS in (3.13) represents absolutely convergent series; this leads to (3.11).

Remark 3.10 Condition (3.11) is equivalent to the condition of combined asymptotic expected additivity from (2.9).

3.3 The Gaussian case

Gaussian processes (GPs) form an instructive example casting light upon the structure of the primary WE rate A_0 : they give an opportunity to assess an impact of ergodicity and asymptotic additivity. Here we list and discuss GP properties in a convenient order. Consider a real double-infinite matrix $\mathbf{C} = (C(i, j) : i, j \in \mathbb{Z})$. Assume that, $\forall m < n$, the $(n-m+1) \times (n-m+1)$ bloc $\mathbf{C}_{m,n} = (C(i, j) : m \leq i, j \leq n)$ gives a (strictly) positive definite matrix. A GP $\mathbf{X} = (X_n : n \in \mathbb{Z})$ with zero mean and covariance matrix \mathbf{C} has a family of PDFs $f_{m,n} = f_{\mathbf{C}_{m,n}}^{\text{No}}$, $m < n$, in \mathbb{R}^{n-m+1} , relative to the Lebesgue measure $d\mathbf{x}_{m,n}$. Here

$$f_{m,n}(\mathbf{x}_{m,n}) = \frac{1}{\left[(2\pi)^{n-m+1} \det \mathbf{C}_{m,n}\right]^{1/2}} \exp\left(-\frac{\mathbf{x}_{m,n}^T \mathbf{C}_{m,n}^{-1} \mathbf{x}_{m,n}}{2}\right), \quad (3.14)$$

$$\mathbf{x}_{m,n}^T = \mathbf{x}_m^n = (x_m, \dots, x_n) \in \mathbb{R}^{n-m+1}.$$

In this section, $\mathbf{x}_{m,n}$ stands for a column- and $\mathbf{x}_{m,n}^T$ for a row-vector. (A similar rule will be applied to random vectors $\mathbf{X}_{m,n}$ and $\mathbf{X}_{m,n}^T$.) When $m = 0$ we write f_n for $f_{0,n}$ and \mathbf{C}_n for $\mathbf{C}_{0,n}$.

If entries $C(i, j)$ have the property $C(i, j) = C(0, j - i)$, process \mathbf{X} is stationary. In this case the spectral measure is a (positive) measure μ on $[-\pi, \pi)$ such that $C(i, j) = \int_{-\pi}^{\pi} \cos[(j - i)s] \mu(ds)$. A stationary GP \mathbf{X} is ergodic iff μ has no atoms. Various forms of regularity (decay of correlation) of GPs have been presented in great detail in [7]. We want to note that in Theoretical and Applied Probability (as well as in Statistics), the basic parameter is, typically, \mathbf{C} . On the other hand, in Mathematical Physics it is usually the family of matrices $\mathbf{C}_{m,n}^{-1}$: their entries $C_{m,n}^{(-1)}(i, j)$ play the role of interaction potentials between sites $m \leq i, j \leq n$ for a system of ‘spins’ $x_m, \dots, x_n \in \mathbb{R}$. In this interpretation, the quadratic form $\frac{1}{2} \mathbf{x}_{m,n}^T \mathbf{C}_{m,n}^{-1} \mathbf{x}_{m,n}$ represents the potential energy of a spin configuration $\mathbf{x}_{m,n}$. In these terms, a Markov GP arises when matrices $\mathbf{C}_{m,n}^{-1}$ are tri-diagonal Jacobi; cf. Eqn (4.29) below. The SE $H(f_{m,n}) = \frac{1}{2} \log [e(2\pi)^n (\det \mathbf{C}_n)] = \frac{1}{2} [n \log (2\pi e) - \text{tr } \mathbf{L}_n]$ where $\mathbf{L}_n = \log \mathbf{C}_n^{-1}$.

Now take $m = 0$. Given a WF $\mathbf{x}_{0,n-1} \in \mathbb{R}^n \mapsto \phi_n(\mathbf{x}_{0,n-1})$, the WI and WE have the form

$$I_{\phi_n}^w(\mathbf{x}_0^{n-1}) = \frac{\log [(2\pi)^n (\det \mathbf{C}_n)]}{2} \phi_n(\mathbf{x}_{0,n-1}) + \frac{\log e}{2} (\mathbf{x}_{0,n-1}^T \mathbf{C}_n^{-1} \mathbf{x}_{0,n-1}) \phi_n(\mathbf{x}_{0,n-1}) \quad (3.15)$$

and

$$\begin{aligned} H_{\phi_n}^w(f_n) &= \frac{1}{2} \log [(2\pi)^n (\det \mathbf{C}_n)] \int_{\mathbb{R}^n} \phi_n(\mathbf{x}_{0,n-1}) f_n(\mathbf{x}_{0,n-1}) d\mathbf{x}_{0,n-1} \\ &\quad + \frac{\log e}{2} \int_{\mathbb{R}^n} (\mathbf{x}_{0,n-1}^T \mathbf{C}_n^{-1} \mathbf{x}_{0,n-1}) \phi_n(\mathbf{x}_{0,n-1}) f_n(\mathbf{x}_{0,n-1}) d\mathbf{x}_{0,n-1} \\ &= \left[H(f_n) - n \frac{\log e}{2} \right] \mathbb{E} \phi_n(\mathbf{X}_{0,n-1}) + \frac{\log e}{2} \mathbb{E} \left[(\mathbf{X}_{0,n-1}^T \mathbf{C}_n^{-1} \mathbf{X}_{0,n-1}) \phi_n(\mathbf{X}_{0,n-1}) \right]. \end{aligned} \quad (3.16)$$

Consequently, a finite rate $h = \lim_{n \rightarrow \infty} \frac{1}{n} H(f_n)$ exists iff

$$\lim_{n \rightarrow \infty} \frac{1}{n} \text{tr } \mathbf{L}_n = -h + \log(2\pi e), \quad (3.17)$$

regardless of ergodicity (and even stationarity) of GP \mathbf{X} . Moreover, under assumption (3.17), we obtain that

$$\frac{H_{\phi_n}^w(f_n) - (\log e) \mathbb{E}[(\mathbf{X}_{0,n-1}^T \mathbf{C}_n^{-1} \mathbf{X}_{0,n-1}) \phi_n(\mathbf{X}_{0,n-1})]/2}{n \mathbb{E} \phi_n(\mathbf{X}_{0,n-1})} \rightarrow h - \frac{\log e}{2} \quad (3.18)$$

for any choice of the WFs ϕ_n such that $\mathbb{E} \phi_n(\mathbf{X}_{0,n-1}) \neq 0$. For an asymptotically additive WF ϕ_n satisfying (2.6) and for a GP obeying (3.17), Eqn (3.18) takes the form

$$\frac{H_{\phi_n}^w(f_n) - (\log e) \alpha n^2 / 2}{\alpha n^2} \rightarrow h - \frac{\log e}{2}.$$

This yields (2.14 i) with $A_0 = \alpha h$, again without using ergodicity/stationarity of \mathbf{X}_0^∞ .

Similarly, (3.15) and (3.17) imply that $\forall \mathbf{x} \in \mathcal{X}$,

$$\frac{I_{\phi_n}^w(\mathbf{x}_0^{n-1}) - (\log e) (\mathbf{x}_{0,n-1}^T \mathbf{C}_n^{-1} \mathbf{x}_{0,n-1}) \phi_n(\mathbf{x}_{0,n-1})/2}{\phi_n(\mathbf{x}_{0,n-1})} \rightarrow h - \frac{\log e}{2} \quad (3.19)$$

for any choice of the WFs ϕ_n such that $\phi_n(\mathbf{x}_{0,n-1}) \neq 0$.

On the other hand, take $\phi_n(\mathbf{x}_{0,n-1}) = \alpha n$ (an additive WF with $\varphi(x) = \alpha$). Then Eqn (3.16) becomes

$$H_{\phi_n}^w(f_n) = \frac{\alpha n}{2} [n \log(2\pi e) - \text{tr } \mathbf{L}_n] = \alpha n H(f_n). \quad (3.20)$$

The asymptotics for the WE $H_{\phi_n}^w(f_n)$ and SE $H(f_n)$ will be determined by a ‘competition’ between the terms in the square brackets (an entropy-energy argument in Mathematical Physics). Viz., take $\mathbf{L}_n = (L_n(i, j), 0 \leq i, j < n)$ and suppose that the diagonal entries decrease to $-\infty$ when j is large (say, $L(j, j) \sim -\log(c + j)$ with a constant $c > 0$ or $\lambda_j \sim e^{(c+j)}$ where $\lambda_0 \leq \lambda_1 \leq \dots \leq \lambda_{n-1}$ are the eigen-values of \mathbf{C}_n). Then the trace $\text{tr } \mathbf{L}_n = \sum_{0 \leq j < n} L_n(j, j)$ will dominate, and the correct scale for the rate of $H_{\phi_n}^w(f_n)$ with

$\phi_n(\mathbf{x}_{0,n-1}) = \alpha n$ will be $\frac{1}{n^2 \log n}$.

The above example can be generalised as follows. Let $\mathbf{A} = (A(i, j) : i, j \in \mathbb{Z})$ be a double-infinite real symmetric matrix (with $A(i, j) = A(j, i)$) and consider, $\forall m < n$, the bloc $\mathbf{A}_{m,n} = (A(i, j) : m \leq i, j \leq n)$. Then set

$$\phi_{m,n}(\mathbf{x}_{m,n}) = \mathbf{x}_{m,n}^T \mathbf{A}_{m,n} \mathbf{x}_{m,n}. \quad (3.21)$$

For $\mathbf{A}_{0,n-1}$ we write \mathbf{A}_n . Pictorially, we try to combine a Gaussian form of the PDFs $f_{m,n}(\mathbf{x}_{m,n})$ with a log-Gaussian form of $\phi_{m,n}(\mathbf{x}_{m,n})$.

Then the expression for the WI $I_{\phi_n}^w(\mathbf{x}_{0,n-1}) = -\phi_n(\mathbf{x}_{0,n-1}) \log f_n(\mathbf{x}_{0,n-1})$ and WE $H_{\phi_n}^w(f_n) = \mathbb{E} I_{\phi_n}^w(\mathbf{X}_{0,n-1})$ become

$$\begin{aligned} I_{\phi_n}^w(\mathbf{x}_{0,n-1}) &= (\mathbf{x}_{0,n-1}^T \mathbf{A}_n \mathbf{x}_{0,n-1}) \\ &\times \left\{ \left[H(f_n) - n \frac{\log e}{2} \right] + (\mathbf{x}_{0,n-1}^T \mathbf{C}_n^{-1} \mathbf{x}_{0,n-1}) \log e \right\} \end{aligned} \quad (3.22)$$

and

$$H_{\phi_n}^w(f_n) = \left[H(f_n) - n \frac{\log e}{2} \right] \mathbb{E} \left(\mathbf{X}_{0,n-1}^T \mathbf{A}_n \mathbf{X}_{0,n-1} \right) + \frac{\log e}{2} \mathbb{E} \left[\left(\mathbf{X}_{0,n-1}^T \mathbf{C}_n^{-1} \mathbf{X}_{0,n-1} \right) \left(\mathbf{X}_{0,n-1}^T \mathbf{A}_n \mathbf{X}_{0,n-1} \right) \right]. \quad (3.23)$$

As before, the analysis of rates for (3.22) and (3.23) can be done by comparing the contributions from different terms.

4 Rates for multiplicative WFs

Multiplicative weighted rates behave differently and require a diverse approach to their studies. To start with, the WI rate in general does not coincide with the corresponding WE rate.

4.1 WI rates

The question of a multiplicative WI rate is relatively simple:

Theorem 4.1 *Given an ergodic RP \mathbf{X} with a probability distribution \mathbb{P} on \mathcal{X} , consider the WI $I_{\phi_n}^w(\mathbf{x}_0^{n-1})$ as defined in (??) and (3.1). Suppose that convergence in (2.11) holds \mathbb{P} -a.s. Then convergence in (2.15 i) holds true \mathbb{P} -a.s., where $\bar{B} = \beta$ and the value β is as in (2.11).*

Proof. The assertion follows immediately from the \mathbb{P} -a.s. convergence in Eqn (2.11). ■

4.2 WE rates. The Markovian case

Passing to multiplicative WE rates, we consider in this paper a relatively simple case where (a) RP \mathbf{X}_0^∞ is a homogeneous MC with a stationary PM/DF $\pi(x)$ and the conditional PM/DF $p(y|x)$ and (b) the WF $\phi_n(\mathbf{x}_0^{n-1})$ is a product: for $x, y \in \mathcal{X}$ and $\mathbf{x}_0^{n-1} = (x_0, \dots, x_{n-1}) \in \mathcal{X}^n$,

$$p(y|x) = \frac{\mathbb{P}(X_k \in dy | X_{k-1} = x)}{\nu(dy)}, \quad \phi_n(\mathbf{x}_0^{n-1}) = \prod_{0 \leq j < n} \varphi(x_j). \quad (4.1)$$

In this sub-section we assume that $\varphi(x) \geq 0$ on \mathcal{X} and adopt some positivity assumptions on $p(y|x)$: there exists $k \geq 0$ such that

$$p^{(k+1)}(y|x) = \int_{\mathcal{X}^k} p(y|u_k) \cdots p(u_1|x) \nu^k(d\mathbf{u}_1^k) > 0. \quad (4.2)$$

As earlier, λ stands for an initial PM/DF on \mathcal{X} . Accordingly, we consider the WE $H_{\phi_n}^w(f_n, \lambda)$ of the form

$$\begin{aligned} H_{\phi_n}^w(f_n, \lambda) &= -\mathbb{E}_\lambda \left\{ \prod_{0 \leq j < n} \varphi(X_j) \log \left[\lambda(X_0) \prod_{1 \leq l < n} p(X_l | X_{l-1}) \right] \right\} \\ &= -\int_{\mathcal{X}^n} \lambda(x_0) \varphi(x_0) \prod_{1 \leq i < n} [p(x_i | x_{i-1}) \varphi(x_i)] \\ &\quad \times \left[\log \lambda(x_0) + \sum_{1 \leq l < n} \log p(x_l | x_{l-1}) \right] \nu^n(d\mathbf{x}_0^{n-1}), \end{aligned} \quad (4.3)$$

and the WE $H_{\phi_n}^w(f_n)$ obtained by replacing λ with π .

The product $\varphi(x_0) \prod_{1 \leq i < n} [p(x_i | x_{i-1}) \varphi(x_i)]$ can of course be written in a symmetric (or dual) formation, as $\prod_{1 \leq i < n} [\varphi(x_{i-1}) p(x_i | x_{i-1})] \varphi(x_{n-1})$. It would lead to an equivalent form of results that follow.

The existence (and a number of properties) of the WER B_0 in (2.15 ii) are related to an integral operator W acting on functions $\mathbf{f} : \mathcal{X} \rightarrow \mathbb{R}$ and connected to the conditional PM/DF $p(y|x)$ and factor $\varphi(x)$ in (4.1). Namely, for $y \in \mathcal{X}$, the value $(W\mathbf{f})(y)$ is defined by

$$(W\mathbf{f})(y) = \int_{\mathcal{X}} W(y, w) \mathbf{f}(w) \nu(dw). \quad (4.4)$$

We also introduce an adjoint/transposed operator W^T with an action $\mathbf{g} \mapsto \mathbf{g}W^T$:

$$(\mathbf{g}W^T)(y) = \int_{\mathcal{X}} \mathbf{g}(w) W(w, y) \nu(dw). \quad (4.5)$$

Here the kernel W given as follows: for $u, v \in \mathcal{X}$,

$$W(u, v) = \varphi(u) p(v|u). \quad (4.6)$$

Remark 4.2 The form of writing the action of the adjoint operator as $\mathbf{g}W^T$ does not have a particular significance but shortens and makes more transparent some relations where W and W^T take part. Viz., we have that

$$\int_{\mathcal{X}} \mathbf{g}(y) (W\mathbf{f})(y) \nu(dy) = \int_{\mathcal{X}} (\mathbf{g}W^T)(y) \mathbf{f}(y) \nu(dy),$$

or, in brief, $\langle \mathbf{g}, W\mathbf{f} \rangle = \langle \mathbf{g}W^T, \mathbf{f} \rangle$ where $\langle \mathbf{g}, \mathbf{f} \rangle = \int_{\mathcal{X}} \mathbf{f}(y) \mathbf{g}(y) \nu(dy)$ is the inner product in the (real) Hilbert space $L_2(\mathcal{X}, \nu)$. Also, it emphasizes analogies with a MC formalism where a transition operator acts on functions while its adjoint (dual) acts on measures.

Pictorially speaking, kernel $W^T(x_{i-1}; x_i)$ represents the factor in the product $\prod_{1 \leq i < n} [p(x_i | x_{i-1}) \varphi(x_{i-1})]$ in (4.3) where variable x_i appears for the first time. Accordingly:

$$\begin{aligned} H_{\phi_n}^w(f_n, \lambda) &= - \int_{\mathcal{X}^n} \lambda(x_0) \left\{ [\log \lambda(x_0)] \prod_{1 \leq i < n} W(x_{i-1}, x_i) \varphi(x_{n-1}) \right. \\ &\quad + \sum_{1 \leq l < n} \prod_{1 \leq i \leq l} W^T(x_{i-1}, x_i) \\ &\quad \left. \times [\log p(x_l | x_{l-1})] \prod_{l < j < n} W(x_{j-1}, x_j) \varphi(x_{n-1}) \right\} \nu^n(d\mathbf{x}_0^{n-1}). \end{aligned} \quad (4.7)$$

We will use the following condition (of the Hilbert–Schmidt type):

$$\int_{\mathcal{X} \times \mathcal{X}} W(x, y) W(y, x) \nu(dx) \nu(dy) < \infty. \quad (4.8)$$

Also, suppose that function

$$(x, y) \in \mathcal{X} \times \mathcal{X} \mapsto p(y|x) |\log p(y|x)| \quad (4.9)$$

is bounded and functions

$$x \mapsto \varphi(x), \quad x \mapsto \lambda(x) \log \lambda(x), \quad x \mapsto \pi(x) \log \pi(x) \quad (4.10)$$

belong to $L_2(\mathcal{X}, \nu)$.

Theorem 4.3 *Assume the stated conditions upon \mathbf{X}_0^∞ , transitions PM/DF $p(y|x)$ and WF ϕ_n . Then Eqn (2.15 ii) holds true, both for $H_{\phi_n}^w(f_n)$ and $H_{\phi_n}^w(f_n; \lambda)$:*

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log H_{\phi_n}^w(f_n, \lambda) = \lim_{n \rightarrow \infty} \frac{1}{n} \log H_{\phi_n}^w(f_n) = B_0. \quad (4.11)$$

Here

$$B_0 = \log \mu \quad (4.12)$$

and $\mu > 0$ is the maximal eigen-value of operator \mathbf{W} in $L_2(\mathcal{X}, \nu)$ coinciding with the norm of \mathbf{W} and \mathbf{W}^T ; cf. (4.4). That is, $\mu = \|\mathbf{W}\| = \|\mathbf{W}^T\|$.

Proof. As follows from the previous formulas, we have the following expressions for the WEs $H_{\phi_n}^w(f_n, \lambda)$ and $H_{\phi_n}^w(f_n)$:

$$\begin{aligned} H_{\phi_n}^w(f_n, \lambda) = & - \int_{\mathcal{X}} [\lambda(x_0) \log \lambda(x_0)] (\mathbf{W}^{n-1} \varphi)(x_0) \nu(dx_0) \\ & - \sum_{1 \leq l < n} \int_{\mathcal{X}^2} \left(\lambda \mathbf{W}^{T^{l-1}} \right) (x_{l-1}) [\varphi(x_{l-1}) p(x_l | x_{l-1}) \\ & \quad \times \log p(x_l | x_{l-1})] (\mathbf{W}^{n-1-l} \varphi)(x_l) \nu^2(dx_{l-1} \times dx_l) \end{aligned} \quad (4.13)$$

and

$$\begin{aligned} H_{\phi_n}^w(f_n) = & - \int_{\mathcal{X}} [\pi(x_0) \log \pi(x_0)] (\mathbf{W}^{n-1} \varphi)(x_0) \nu(dx_0) \\ & - \sum_{1 \leq l < n} \int_{\mathcal{X}^2} \left(\pi \mathbf{W}^{T^{l-1}} \right) (x_{l-1}) [\varphi(x_{l-1}) p(x_l | x_{l-1}) \\ & \quad \times \log p(x_l | x_{l-1})] (\mathbf{W}^{n-1-l} \varphi)(x_l) \nu^2(dx_{l-1} \times dx_l). \end{aligned} \quad (4.14)$$

Re-write (4.13) and (4.14) by omitting unnecessary references to l :

$$\begin{aligned} H_{\phi_n}^w(f_n, \lambda) = & - \int_{\mathcal{X}} [\lambda(x) \log \lambda(x)] (\mathbf{W}^{n-1} \varphi)(x) \nu(dx) \\ & - \sum_{1 \leq l < n} \int_{\mathcal{X}^2} \left(\lambda \mathbf{W}^{T^{l-1}} \right) (x) [\varphi(x) p(y|x) \\ & \quad \times \log p(y|x)] (\mathbf{W}^{n-1-l} \varphi)(y) \nu^2(dx \times dy) \end{aligned} \quad (4.15)$$

and

$$\begin{aligned}
H_{\phi_n}^w(f_n) &= - \int [\pi(x) \log \pi(x)] (W^{n-1}\varphi)(x) \nu(dx) \\
&\quad - \sum_{k \leq l < n} \int_{\mathcal{X}^2} \left(\pi W^{T^{l-1}} \right) (x) [\varphi(x) p(y|x) \\
&\quad \times \log p(y|x)] (W^{n-1-l}\varphi)(y) \nu^2(dx \times dy).
\end{aligned} \tag{4.16}$$

At this point we use the *Krein–Rutman theorem* for linear operators preserving the cone of positive functions, which generalizes the Perron–Frobenius theorem for non-negative matrices. The form of the theorem below is a combination of [9], Proposition β , P. 76, and Proposition β' , P. 77. See also [5], Theorem 19.2.

Theorem (Krein–Rutman). *Suppose that \mathcal{Y} is a Polish space and ϖ is a Borel measure on \mathcal{Y} . Assume a non-negative continuous kernel $K(x, y)$ satisfies the condition: \exists an integer $k \geq 0$ such that the iterated kernel satisfies the positivity condition:*

$$K^{(k+1)}(x, y) = \int_{\mathcal{X}^k} K(x, u_1) K(u_1, u_2) \cdots K(u_k, y) \prod_{1 \leq j \leq k} \varpi(du_j) \geq \theta(y) > 0, \quad x, y \in \mathcal{Y}.$$

Consider mutually adjoint integral operators K and K^T in the Hilbert space $L_2(\mathcal{Y}, \varpi)$:

$$K\underline{v}(y) = \int_{\mathcal{Y}} K(y, v) \underline{v}(u) \varpi(du), \quad \underline{v}K^T(y) = \int_{\mathcal{Y}} \underline{v}(u) K(u, y) \varpi(du) \tag{4.17}$$

and assume operators K and K^T are compact. The following assertions hold true. (i) The norm $\|K\| = \|K^T\| := \kappa \in (0, \infty)$ is an eigen-value of K and K^T of multiplicity one, and the corresponding eigen-functions $\underline{\Phi}$ and $\underline{\Psi}$ are strictly positive on \mathcal{X} :

$$K\underline{\Phi} = \kappa\underline{\Phi}, \quad \underline{\Psi}K^T = \kappa\underline{\Psi}; \quad \underline{\Phi}, \underline{\Psi} > 0.$$

(ii) Operators K and K^T have the following contraction properties. Assume that $\underline{\Phi}$ and $\underline{\Psi}$ are chosen so that $\langle \underline{\Phi}, \underline{\Psi} \rangle = 1$. There exists $\delta \in (0, 1)$ such that \forall function $\underline{v} \in L_2(\mathcal{Y}, \varpi)$ with $\|\underline{v}\|^2 = \langle \underline{v}, \underline{v} \rangle = 1$, functions $K^n \underline{v}$ and $\underline{v}K^{T^n}$ have the following asymptotics:

$$\frac{K^n \underline{v}}{\kappa^n} = \langle \underline{v}, \underline{\Phi} \rangle \underline{\Psi} + \underline{Q}_n, \quad \frac{\underline{v}K^{T^n}}{\kappa^n} = \langle \underline{v}, \underline{\Psi} \rangle \underline{\Phi} + \underline{R}_n. \tag{4.18}$$

Here $\langle \cdot, \cdot \rangle$ stands for the scalar product in $L_2(\mathcal{Y}, \varpi)$ and the norma of vectors $\underline{Q}_n, \underline{R}_n$ are exponentially decreasing:

$$\|\underline{Q}_n\|, \|\underline{R}_n\| \leq (1 - \delta)^n.$$

We are going to apply the Krein–Rutman (KR) theorem in our situation. By using the notation $\langle \cdot, \cdot \rangle$ and $\|\cdot\|$ for the scalar product and the norm in $L_2(\mathcal{X}, \nu)$, we can re-write Eqns (4.15) and (4.16):

$$\begin{aligned}
H_{\phi_n}^w(f_n, \lambda) &= - \left\{ \mu \langle \underline{\Psi}, \varphi \rangle \langle \underline{\Phi}, \lambda \log \lambda \rangle + (n-2) \langle \underline{\Psi}, \varphi \rangle \langle \underline{\Phi}, \lambda \rangle \right. \\
&\quad \times \int_{\mathcal{X}^2} \underline{\Phi}(y') \underline{\Psi}(y) [\varphi(y) p(y'|y) \log p(y'|y)] \nu^2(dy \times dy') \Big\} \mu^{n-2} + O((1-\delta)^n), \\
H_{\phi_n}^w(f_n) &= - \left\{ \mu \langle \underline{\Psi}, \varphi \rangle \langle \underline{\Phi}, \pi \log \pi \rangle + (n-2) \langle \underline{\Psi}, \varphi \rangle \langle \underline{\Phi}, \pi \rangle \right. \\
&\quad \times \int_{\mathcal{X}^2} \underline{\Phi}(y') \underline{\Psi}(y) [\varphi(y) p(y'|y) \log p(y'|y)] \nu^2(dy \times dy') \Big\} \mu^{n-2} + O((1-\delta)^n).
\end{aligned} \tag{4.19}$$

This yields that $H_{\phi_n}^w(f_n, \lambda) \asymp \mu^n$ and $H_{\phi_n}^w(f_n) \asymp \mu^n$, or, formally,

$$\frac{1}{n} \log H_{\phi_n}^w(f_n, \lambda), \frac{1}{n} \log H_{\phi_n}^w(f_n) \rightarrow \log \mu. \quad (4.20)$$

Here $\mu = \|W\| = \|W^T\|$ is the positive eigen-value of operators W and W^T , $\underline{\Phi}$ and $\underline{\Psi}$ are the positive eigen-vectors of W and W^T , respectively, as in the KR theorem. The value $\delta \in (0, 1)$ represents a spectral gap for W and W^T . ■

We will call μ as a KR eigen-value of operator W .

Remark 4.4 The expressions in the curled brackets in (4.19) do not play a role in determining the prime rate B_0 . However, they when we discuss the secondary rate B_1 . Cf. Eqns (4.32), (4.33) below.

Remark 4.5 An assertion similar to Theorem 6 can be proven for a general initial distribution λ (not necessarily absolutely continuous with respect to ν).

Remark 4.6 The Markovian assumption adopted in Theorem 6 can be relaxed without a problem to the case of a Markov chain of order k . Further steps require an extension of this techniques. See Remark 4.10 below.

The relations (4.18) in the KR theorem helps with identifying not only the value B_0 but also B_1 arising from a generalisation of (2.4) for MCs \mathbf{X}_0^∞ of order k . More precisely, with the help of (4.19) we can establish

Theorem 4.7 *Under assumptions of Theorem 4.3,*

$$\begin{aligned} \lim_{n \rightarrow \infty} \frac{H_{\phi_n}^w(f_n)}{n\mu^n} &= \lim_{n \rightarrow \infty} \frac{H_{\phi_n}^w(f_n, \lambda)}{n\mu^n} = -\frac{1}{\mu^2} \langle \underline{\Psi}, \varphi \rangle \langle \underline{\Phi}, \pi \rangle \\ &\times \int_{\mathcal{X}^2} \underline{\Phi}(x) \underline{\Psi}(y) [\varphi(x) p(y|x) \log p(y|x)] \nu^2(dx \times dy). \end{aligned} \quad (4.21)$$

It is instructive to consider a stationary and ergodic MC, with distribution $\tilde{\mathbb{P}}$ on \mathcal{X} which is constructed as follows. The conditional and equilibrium PM/DFs for this MC, $\tilde{p}(y|x) = \frac{\tilde{\mathbb{P}}(X_k \in dy | X_{k-1} = x)}{\nu(dy)}$ and $\tilde{\pi}(x) = \frac{\tilde{\mathbb{P}}(X_k \in dx)}{\nu(dx)}$, for $x, y \in \mathcal{X}$, are given by

$$\tilde{p}(y|x) = \frac{W(x, y) \Phi(y)}{\mu \Phi(x)}, \quad \tilde{\pi}(\mathbf{x}_0^{k-1}) = \Psi(\mathbf{x}_0^{k-1}) \Phi(\mathbf{x}_0^{k-1}),$$

assuming the normalization $\langle \Psi, \Phi \rangle = \int_{\mathcal{X}} \Psi(x) \Phi(x) = 1$. The n -string PM/DF $\tilde{f}_n(\mathbf{x}_0^{n-1}) =$

$\tilde{\pi}(x_0) \prod_{j=1}^{n-1} \tilde{p}(x_j | x_{j-1})$ generated by $\tilde{\mathbb{P}}$ has the form

$$\tilde{p}_n(\mathbf{x}_0^{n-1}) = \Psi(x_0) \prod_{j=k}^{n-1} p(x_j | x_{j-k-1}) \Phi(x_{n-1}).$$

The asymptotic behaviour of the WE $H_{\phi_n}^w(f_n)$ for a multiplicative WF ϕ_n is closely related to properties important in Mathematical Physics and the theory of Dynamical systems. In this regard, we provide here the following assertion which is known as the variational principle for the pressure, entropy and energy. In our context, for a Markov chain \mathbf{X}_0^∞ under the above assumptions, these concepts can be introduced in a variety of forms. Viz., for the metric pressure we can write:

$$\begin{aligned} B_0 &= \log \mu = \lim_{n \rightarrow \infty} \frac{1}{n} \log \Xi_n \\ \text{where } \Xi_n &= \int_{\mathcal{X}^n} \pi(x_0) \prod_{1 \leq j < n} W(x_{j-1}, x_j) \varphi(x_{n-1}) \nu^n(\mathbf{x}_0^{n-1}) \end{aligned} \quad (4.22)$$

and introduce a PM/DF \bar{p}_n :

$$\bar{p}_n(\mathbf{x}_0^{n-1}) = \frac{1}{\Xi_n} \pi(x_0) \prod_{1 \leq j < n} W(x_{j-1}, x_j) \varphi(x_{n-1}) \nu^n(\mathbf{x}_0^{n-1}), \quad (4.23)$$

with $\int_{\mathcal{X}^n} \bar{p}_n(\mathbf{x}_0^{n-1}) \nu^n(d\mathbf{x}_0^{n-1}) = 1$.

Note that

$$\frac{\tilde{p}_n(\mathbf{x}_0^{n-1})}{\bar{p}_n(\mathbf{x}_0^{n-1})} = \frac{\Xi_n \Psi(x_0)}{\mu^{n-1} \Phi(x_{n-1})}$$

and therefore

$$\lim_{n \rightarrow \infty} \frac{1}{n} \int_{\mathcal{X}^n} \log \frac{\tilde{p}_n(\mathbf{x}_0^{n-1})}{\bar{p}_n(\mathbf{x}_0^{n-1})} \tilde{p}_n(\mathbf{x}_0^{n-1}) \nu^n(d\mathbf{x}_0^{n-1}) = \lim_{n \rightarrow \infty} \frac{\log \Xi_n}{n} - \log \mu = 0. \quad (4.24)$$

Theorem 4.8 *Assume the conditions of Theorem 6 for the Markov chain \mathbf{X}_0^∞ with distribution \mathbb{P} . Let \mathbb{Q} be a probability distribution on \mathcal{X}_0^∞ , with $\mathbb{Q}(\mathbf{X}_0^{n-1} \in d\mathbf{x}_0^{n-1}) \prec \nu^n(d\mathbf{x}_0^{n-1})$ and $q^{(n)}(\mathbf{x}_0^{n-1}) = \frac{\mathbb{Q}(\mathbf{X}_0^{n-1} \in d\mathbf{x}_0^{n-1})}{\nu^n(d\mathbf{x}_0^{n-1})}$, for which there exist finite rates of the SE and the log of the kernel W :*

$$h(\mathbb{Q}) = \lim_{n \rightarrow \infty} \frac{-1}{n} \mathbb{E}_{\mathbb{Q}} \log q_n(\mathbf{X}_0^{n-1}), \quad L(\varphi, \mathbb{Q}) = \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{j=k}^{n-1} \mathbb{E}_{\mathbb{Q}} \log W(X_{j-1}, X_j). \quad (4.25)$$

Then the quantity $B_0 = \log \mu$ calculated for \mathbb{P} satisfies the inequality

$$h(\mathbb{Q}) + L(\varphi, \mathbb{Q}) \leq B_0. \quad (4.26)$$

For $\mathbb{Q} = \tilde{\mathbb{P}}$, we have equality. Furthermore, suppose that for a stationary and ergodic \mathbb{Q} we have equality in (4.26). Then $\mathbb{Q} = \mathbb{P}$.

Proof. The core of the argument used in the proof below is well-known in the literature in Mathematical Physics and the theory of Dynamical systems. We write

$$\begin{aligned} 0 &\leq \int_{\mathcal{X}^n} \log \frac{q_n(\mathbf{x}_0^{n-1})}{\bar{p}_n(\mathbf{x}_0^{n-1})} q_n(\mathbf{x}_0^{n-1}) \nu^n(d\mathbf{x}_0^{n-1}) \text{ (by Gibbs' inequality)} \\ &= \mathbb{E}_{\mathbb{Q}} \log q_n(\mathbf{X}_0^{n-1}) - \mathbb{E}_{\mathbb{Q}} \log \prod_{j=k}^{n-1} W(X_{j-1}, X_j) + \log \Xi_n. \end{aligned} \quad (4.27)$$

Dividing by n and passing to the limit yields (4.26).

Now, for $\mathbb{Q} = \widetilde{\mathbb{P}}$, we use (4.24); this yields equality in (4.26).

Finally, let \mathbb{Q} be a stationary process for which $\mathbf{h}(\mathbb{Q}) + \mathbf{L}(\psi, \mathbb{Q}) = \mathbf{B}_0$. It suffices to check that \forall given positive integer m , we have $\mathbb{E}_{\mathbb{Q}}\mathbf{g}(\mathbf{X}) = \mathbb{E}_{\mathbb{P}^*}\mathbf{g}(\mathbf{X})$ for any measurable and bounded function \mathbf{g} depending on \mathbf{x}_0^{m-1} . From (4.27) and (4.24) we deduce that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\mathbb{Q}} \log \frac{q_n(\mathbf{X}_0^{n-1})}{\bar{p}_n(\mathbf{X}_0^{n-1})} = 0 \quad \text{and hence} \quad \lim_{n \rightarrow \infty} \frac{1}{n} \mathbb{E}_{\mathbb{Q}} \log \frac{q_n(\mathbf{X}_0^{n-1})}{p_n^*(\mathbf{X}_0^{n-1})} = 0.$$

Then for n large enough the ratio $f_{0,n-1}^*(\mathbf{x}_0^{n-1}) := \frac{q_n(\mathbf{x}_0^{n-1})}{p_n^*(\mathbf{x}_0^{n-1})} < \infty$ whenever $p_n^*(\mathbf{x}_0^{n-1}) > 0$. So, $f_{0,n-1}^*$ yields the Radon–Nikodym derivative. Moreover, setting $f_{m,n-1}^*(\mathbf{x}_m^{n-1}) = \mathbb{E}_{\mathbb{P}^*} [f_{0,n-1}(\mathbf{X}_0^{n-1}) | \mathbf{X}_m^{n-1} = \mathbf{x}_m^{n-1}]$, we have that

$$\lim_{n \rightarrow \infty} \mathbb{E}_{\mathbb{P}^*} |f_{0,n-1}^*(\mathbf{X}_0^{n-1}) - f_{m,n-1}^*(\mathbf{X}_m^{n-1})| = 0.$$

Then writing:

$$\begin{aligned} \mathbb{E}_{\mathbb{Q}}\mathbf{g}(\mathbf{X}_0^{m-1}) - \mathbb{E}_{\mathbb{P}^*}\mathbf{g}(\mathbf{X}_0^{m-1}) &= \mathbb{E}_{\mathbb{Q}} [\mathbf{g}(\mathbf{X}_0^{m-1}) - \mathbb{E}_{\mathbb{P}^*}\mathbf{g}(\mathbf{X}_0^{m-1})] \\ &= \mathbb{E}_{\mathbb{P}^*} [f_{0,n-1}^*(\mathbf{X}_0^{n-1})\mathbf{g}(\mathbf{X}_0^{m-1}) - f_{m,n-1}^*(\mathbf{X}_m^{n-1})\mathbf{g}(\mathbf{X}_0^{m-1})] \\ &= \mathbb{E}_{\mathbb{P}^*} \{ [f_{0,n-1}^*(\mathbf{X}_0^{n-1}) - f_{m,n-1}^*(\mathbf{X}_m^{n-1})]\mathbf{g}(\mathbf{X}_0^{m-1}) \} \end{aligned}$$

yields the desired result. ■

Example 4.9 As an example where the above assumptions are fulfilled, consider the case where $\mathcal{X} = \mathbb{Z}_+ = \{0, 1, \dots\}$, and ν is the counting measure ($\nu(i) = 1, i \in \mathbb{Z}_+$). The proposed transition PMF is

$$p(y|x) = [1 - e^{-(x+1)}]e^{-(x+1)y}, \quad x, y \in \mathbb{Z}_+,$$

with the stationary PMF

$$\pi(x) = \Xi^{-1}e^{-x}[1 - e^{-(x+1)}], \quad x \in \mathbb{Z}_+, \quad \text{where} \quad \Xi = \sum_{u \in \mathbb{Z}_+} e^{-u}[1 - e^{-(u+1)}].$$

Conditions (4.8), (4.9) and (4.10) will be fulfilled when we choose $\varphi \in \ell_2(\mathbb{Z}_+)$.

In a continuous setting: let $\mathcal{X} = \mathbb{R}_+$, with ν being a Lebesgue measure. The transition PDF is given by

$$p(y|x) = (x+1)e^{-(x+1)y}, \quad x, y \in \mathbb{R}_+,$$

with the stationary PDF

$$\pi(x) = \Xi^{-1} \frac{e^{-x}}{x+1}, \quad x \in \mathbb{R}_+, \quad \text{where} \quad \Xi = \int_0^\infty \frac{e^{-u}}{u+1} du.$$

Here conditions (4.8), (4.9) and (4.10) will be fulfilled when we choose $\varphi \in L_2(\mathbb{R}_+, \nu)$.

Remark 4.10 In order to move beyond Markovian assumptions upon process $\mathbf{X} = \{X_i : i \in \mathbb{Z}\}$, one has to introduce conditions controlling conditional PM/DF

$$p(y|\mathbf{x}_{-\infty}^0) = \frac{\mathbb{P}(X_1 \in dy | \mathbf{X}_{-\infty}^0 = \mathbf{x}_{-\infty}^0)}{\nu(dy)}.$$

At present, a sufficiently complete theory exists for the case of a compact space \mathcal{X} , based on the theory of Gibbs measures. A standard reference here is [11]. See also [12], Ch. 5.6, [13], Ch. 5, [6], Ch. 8.3 and the relevant bibliography therein. Extensions to non-compact cases require further work; we intend to return to this topic in forthcoming papers. Among related papers, Refs [19], [20] may be of some interest here.

4.3 WE rates for Gaussian processes

As before, it is instructive to discuss the Gaussian case. A well-known model of a (real-valued) Markov GP $\mathbf{X}_0^\infty = (X_0, X_1, \dots)$ is described via a stochastic equation

$$X_{n+1} = \alpha X_n + Z_{n+1}, \quad n \geq 0. \quad (4.28)$$

Cf. [2]. Here $\{Z_n, n \in \mathbb{Z}\}$ is a sequence of IID random variables, where $Z_n \sim N(0, 1)$ has $\mathbb{E}Z_n = 0$ and $\text{Var } Z_n = 1$. (A general case $Z_n \sim N(0, \sigma^2)$ does not add a serious novelty.) The transition PDF $p(x, y)$ has the form $p(x, y) = \frac{e^{-(y-\alpha x)^2/2}}{\sqrt{2\pi}}$, $x, y \in \mathbb{R}$. The constant α will be taken from the interval $(-1, 1)$, with $|\alpha| < 1$. To obtain a stationary process, we take $X_0 \sim N(0, c)$ where $c = \frac{1}{1-\alpha^2}$. This results in the (strong) solution $X_n = \sum_{l \geq 0} \alpha^l Z_{n-l}$, $n \in \mathbb{Z}$ (the series converge almost surely) and defines process \mathbf{X} with

probability measure \mathbb{P} on $\mathbb{R}^{\mathbb{Z}}$ and expectation \mathbb{E} . The equilibrium PDF is $\pi(x) = \frac{e^{-x^2/(2c)}}{\sqrt{2\pi c}}$, $x \in \mathbb{R}$. Given $n > 2$, the joint PDF $f_n(\mathbf{x}_0^{n-1}) = \pi(x_0) \prod_{1 \leq j < n} p(x_{j-1}, x_j)$ for \mathbf{X}_0^{n-1} has the form

$$\begin{aligned} f_n(\mathbf{x}_0^{n-1}) &= \frac{\sqrt{1-\alpha^2}}{(2\pi)^{n/2}} \exp \left(-\frac{1}{2} \left\{ x_0^2 - \alpha x_0 x_1 \right. \right. \\ &\quad \left. \left. + \sum_{1 \leq j \leq n-2} \left[-\alpha(x_{j-1} + x_{j+1})x_j + (1+\alpha^2)x_j^2 \right] - \alpha x_{n-2}x_{n-1} + x_{n-1}^2 \right\} \right) \\ &= \frac{\sqrt{1-\alpha^2}}{(2\pi)^{n/2}} \exp \left[-\frac{x_0^2}{2} + \alpha x_0 x_1 - (1+\alpha^2)\frac{x_1^2}{2} + \alpha x_1 x_2 - (1+\alpha^2)\frac{x_2^2}{2} \right. \\ &\quad \left. + \dots - (1+\alpha^2)\frac{x_{n-2}^2}{2} + \alpha x_{n-2}x_{n-1} - \frac{x_{n-1}^2}{2} \right], \quad \mathbf{x}_0^{n-1} = (x_0, \dots, x_{n-1}) \in \mathbb{R}^n. \end{aligned} \quad (4.29)$$

Thus, $f_n \sim N(\mathbf{0}, \mathbf{C}_n)$ where \mathbf{C}_n is the inverse of a Jacobi $n \times n$ matrix

$$\mathbf{C}_n^{-1} = \begin{pmatrix} 1 & -\alpha & 0 & \dots & 0 & 0 \\ -\alpha & 1+\alpha^2 & -\alpha & \dots & 0 & 0 \\ 0 & -\alpha & 1+\alpha^2 & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & 1+\alpha^2 & -\alpha \\ 0 & 0 & 0 & \dots & -\alpha & 1 \end{pmatrix};$$

cf. (3.14). Assume that $\phi_n(\mathbf{x}_0^{n-1}) = \prod_{0 \leq j < n} \varphi(x_j)$ (a special case of (??), with $k = 1$). The WE $H_{\phi_n}^w(f_n) = -\mathbb{E}\phi_n(\mathbf{X}_0^{n-1}) \log f_n(\mathbf{X}_0^{n-1})$ takes the form

$$H_{\phi_n}^w(f_n) = \frac{1}{2} \mathbb{E} \left\{ \prod_{0 \leq j < n} \varphi(X_j) \left[X_0^2 - 2\alpha X_0 X_1 + (1 + \alpha^2) X_1^2 - 2\alpha X_1 X_2 + (1 + \alpha^2) X_1^2 - \dots + (1 + \alpha^2) X_{n-2}^2 - 2\alpha X_{n-2} X_{n-1} + X_{n-1}^2 - 2 \log \frac{\sqrt{1 - \alpha^2}}{(2\pi)^{n/2}} \right] \right\}. \quad (4.30)$$

Discarding border terms (and omitting the factor $1/2$), the bulk structure of $H_{\phi_n}^w(f_n)$ is represented by the sum

$$\sum_{1 \leq l \leq n-2} \mathbb{E} \left\{ \left[(1 + \alpha^2) X_l^2 - 2\alpha X_l X_{l+1} + \log(2\pi) \right] \prod_{0 \leq j < n} \varphi(X_j) \right\}.$$

For a value $1 < l < n - 1$ away from 1 and n , the corresponding summand admits the form

$$\begin{aligned} & \mathbb{E} \left\{ \left[(1 + \alpha^2) X_l^2 - 2\alpha X_l X_{l+1} + \log(2\pi) \right] \prod_{0 \leq j < n} \varphi(X_j) \right\} \\ &= \frac{\sqrt{1 - \alpha^2}}{(2\pi)^{n/2}} \int_{\mathbb{R}^n} e^{-x_0^2(1 - \alpha^2)/2} \varphi(x_0) \prod_{1 \leq j < n} e^{-(x_j - \alpha x_{j-1})^2/2} \varphi(x_j) \\ & \quad \times \left[(1 + \alpha^2) x_l^2 - 2\alpha x_l x_{l+1} + \log(2\pi) \right] dx_0 \dots dx_{n-1}. \end{aligned} \quad (4.31)$$

Following the spirit of the Krein–Rutman theorem we represent (4.31) as

$$\begin{aligned} & \int_{\mathbb{R} \times \mathbb{R}} (\mathbf{W}^{l-1} \varphi_0^*)(x) \left[(1 + \alpha^2) x^2 - 2\alpha x y + \log(2\pi) \right] \frac{e^{-(y - \alpha x)^2/2}}{\sqrt{2\pi}} (\mathbf{W}^{n-l-2} \varphi_1^*)(y) dx dy \\ &= \mu^{n-3} \left\{ \langle \varphi_0^*, \underline{\Psi} \rangle \langle \varphi_1^*, \underline{\Phi} \rangle \int_{\mathbb{R} \times \mathbb{R}} \underline{\Phi}(x) \underline{\Psi}(y) \right. \\ & \quad \times \left. \left[(1 + \alpha^2) y^2 - 2\alpha x y + \log(2\pi) \right] \frac{e^{-(y - \alpha x)^2/2}}{\sqrt{2\pi}} dx dy + O((1 - \delta)^{n-3}) \right\}. \end{aligned} \quad (4.32)$$

As before, $\mu > 0$ is the principal eigen-value of operator \mathbf{W} in $L_2(\mathbb{R})$, given by

$$\mathbf{W}\mathbf{f}(y) = \int_{\mathbb{R}} W(y, u) \mathbf{f}(u) du \quad \text{where} \quad W(y, u) = \varphi(y) \frac{\exp \left[-(y - \alpha u)^2/2 \right]}{\sqrt{2\pi}}.$$

Next, $\underline{\Phi}$ and $\underline{\Phi}^*$ are the corresponding positive eigen-functions of \mathbf{W} and its adjoint \mathbf{W}^T , with $\mathbf{W}\underline{\Phi} = \mu \underline{\Phi}$, $\underline{\Psi} \mathbf{W} = \mu \underline{\Psi}$, $\mu = \|\mathbf{W}\| = \|\mathbf{W}^T\|$. Finally,

$$\varphi_0^*(x) = \frac{\sqrt{1 - \alpha^2}}{\sqrt{2\pi}} e^{-x^2(1 - \alpha^2)/2} \varphi(x), \quad \varphi_1^*(y) = \frac{1}{\sqrt{2\pi}} \int_{\mathbb{R}} e^{-(z - \alpha y)^2/2} \varphi(z) dz, \quad x, y \in \mathbb{R}. \quad (4.33)$$

Assuming suitable conditions on one-step WF φ , this leads to Theorems 4.3 and 4.7.

Remark 4.11 The WE rate for a multiplicative WF can be interpreted as a metric pressure, a concept proved to be useful in the Dynamical system theory. The next step is to introduce a topological pressure, along with its specific case, topological entropy. See [21], Ch. 9.

A simple example of a topological entropy and pressure in our context is as follows. Let $\mathcal{X} = \mathbb{R}$ and $\nu(dx) = p(x)dx$ where $p(x) = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$. Fix a number $a > 0$ and consider the set $\mathcal{A} \subset \mathcal{X}$:

$$\mathcal{A} = \{\mathbf{x} = (x_i : i \in \mathbb{Z}) : |x_i - x_{i+1}| > a \ \forall i \in \mathbb{Z}\}.$$

Define the topological entropy $h^{\text{top}}(\mathcal{A}, \nu)$ by

$$h^{\text{top}}(\mathcal{A}, \nu) = \lim_{n \rightarrow \infty} \frac{1}{n} \log \nu^n(\mathcal{A}_0^{n-1}).$$

Here

$$\mathcal{A}_0^{n-1} = \left\{ \mathbf{x}_0^{n-1} \in \mathcal{X}^n : \mathbf{x}_0^{n-1} = (x_0, \dots, x_{n-1}) : |x_i - x_{i-1}| > a \ \forall 1 \leq i \leq n-1 \right\}.$$

Then $h^{\text{top}}(\mathcal{A}, \nu) = \log \mu$ where μ is the KR eigen-value for operator W in $L_2(\mathbb{R})$ given by

$$(Wg)(x) = \int_{\mathbb{R}} W(x, y)g(y)dy \quad \text{with} \quad W(x, y) = p(x)\mathbf{1}(|x - y| > a).$$

In fact, Theorem 4.3 is applicable here. For the second iteration kernel $W^{(2)}(x, y) = \int_{\mathbb{R}} W(x, u)W(u, y)du$ we have

$$W^{(2)}(x, y) = p(x) \int_{\mathbb{R}} p(u)\mathbf{1}(|u - x| > a, |u - y| > a)du \geq cp(x)$$

where $c = \int \mathbf{1}(|u| > 2a)du$. This implies assumption (4.2) (with $k = 1$). The Hilbert–Schmidt type condition (4.8) is also fulfilled:

$$\int_{\mathbb{R}} W(x, y)W(y, x)dxdy = \int_{\mathbb{R}} p(x)p(y)\mathbf{1}(|x - y| > a)dxdy < 1.$$

At the same time, if we set $\nu_0(dx) = dx$ then $\log \mu$ can be interpreted as the topological pressure $\mathcal{P}^{\text{top}}(\mathcal{A}, \chi, \nu_0)$ for set \mathcal{A} , function $\chi = \ln p$ and reference measure ν_0 :

$$\mathcal{P}^{\text{top}}(\mathcal{A}, \chi, \nu_0) = \lim_{n \rightarrow \infty} \frac{1}{n} \ln \int_{\mathcal{A}_0^{n-1}} \exp \left[\sum_{i=0}^{n-1} \chi(x_i) \right] \nu_0(dx_0) \cdots \nu_0(dx_{n-1}).$$

These connections are worth of further explorations.

5 Rates for multiplicative Gaussian WFs

In this section we focus on rates for Gaussian RPs and WFs. Recall, the SI and SE for a Gaussian PDF $f_{m,n} = f_{\mathbf{C}_{m,n}}^{\text{No}}$ are given by

$$I(\mathbf{x}_{m,n}, f_{m,n}) = \frac{1}{2} \left\{ \log \left[(2\pi)^{n-m+1} \det \mathbf{C}_{m,n} \right] + \mathbf{x}_{m,n}^T \mathbf{C}_{m,n}^{-1} \mathbf{x}_{m,n} \log e \right\} \quad (5.1)$$

and

$$H(f_{m,n}) = \frac{1}{2} \log \left[(2\pi e)^{n-m+1} \det \mathbf{C}_{m,n} \right] = \frac{1}{2} \left[(n-m+1) \log(2\pi e) - \text{tr } \mathbf{L}_{m,n} \right] \quad (5.2)$$

where $\mathbf{L}_{m,n} = \log \mathbf{C}_{m,n}^{-1}$. As before, for $m = 0$, we set: $\mathbf{C}_{0,n-1} = \mathbf{C}_n$ and write f_n for $f_{\mathbf{C}_n}^{\text{No}}$. (A similar agreement will be in place for other matrices/functions emerging below.) We can write $I(\mathbf{x}_{0,n-1}, f_n) = \frac{1}{2} \left[H(f_n) - n \log e \right]$.

First, a simple example. Suppose we take $\phi_n(\mathbf{x}_{0,n-1}) = e^{bn}$ where b is a constant, real or complex. (A special case of a multiplicative WF with $\varphi(x) = b$; cf. (??).) With $\mathbf{L}_n = (L_n(i, j), 0 \leq i, j < n)$, Eqn (3.16) becomes

$$H_{\phi_n}^w(f_n) = \frac{e^{bn}}{2} \left[n \log(2\pi e) - \sum_{0 \leq j < n} L_n(j, j) \right] = e^{bn} H(f_n). \quad (5.3)$$

Assume that $-\frac{1}{n} \sum_{0 \leq j < n} L_n(j, j)$ converges to a value $a \in \mathbb{R}$ as $n \rightarrow \infty$. Then $\frac{1}{n} H(f_n) = \frac{1}{n} \left[n \log(2\pi e) - \text{tr } \mathbf{L}_n \right] \rightarrow \log(2\pi e) + a$. Hence, we obtain $\frac{1}{n} H_{\phi_n}^w(f_n) \asymp e^{bn} [\log(2\pi e) + a]$; if $\log(2\pi e) + a \neq 0$, it implies that

$$\lim_{n \rightarrow \infty} \frac{1}{n} \log \frac{1}{n} H_{\phi_n}^w(f_n) = b.$$

In general, the rate of growth/decay of $H_{\phi_n}^w(f_n)$ is determined by that of $\text{tr } \mathbf{L}_n$.

Next, consider an WF $\mathbf{x}_{m,n} \mapsto \phi_{m,n}(\mathbf{x}_{m,n})$ of the following form. Let $\mathbf{A} = (A(i, j) : i, j \in \mathbb{Z})$ be a double-infinite real symmetric matrix (with $A(i, j) = A(j, i)$) and assume that, $\forall m < n$, the bloc $\mathbf{A}_{m,n} = (A(i, j) : m \leq i, j \leq n)$ is such that matrix $\mathbf{C}_{m,n}^{-1} - \mathbf{A}_{m,n}$ is (strictly) positive definite. Then choose a real double-infinite sequence $\mathbf{t} = (t_n, n \in \mathbb{Z})$ and set

$$\phi_{m,n}(\mathbf{x}_{m,n}) = \exp \left[\mathbf{x}_{m,n}^T (\mathbf{C}_{m,n}^{-1} - \mathbf{A}_{m,n}) \mathbf{t}_{m,n} + \frac{1}{2} \mathbf{x}_{m,n}^T \mathbf{A}_{m,n} \mathbf{x}_{m,n} \right], \quad (5.4)$$

where column-vectors $\mathbf{t}_{m,n} = (t_i : m \leq i \leq n)$, $\mathbf{x}_{m,n} = (x_i : m \leq i \leq n) \in \mathbb{R}^{n-m+1}$.

Then the WI $I_{\phi_{m,n}}^w(\mathbf{x}_{m,n}, f_{m,n}) := -\phi_{m,n}(\mathbf{x}_{m,n}) \log f_{\mathbf{C}_{m,n}}^{\text{No}}(\mathbf{x}_{m,n})$ becomes

$$I_{\phi_{m,n}}^w(\mathbf{x}_{m,n}, f_{m,n}) = \frac{1}{2} \left\{ \log \left[(2\pi)^{n-m+1} \det \mathbf{C}_{m,n} \right] + \mathbf{x}_{m,n}^T \mathbf{C}_{m,n}^{-1} \mathbf{x}_{m,n} \log e \right\} \\ \times \exp \left[\mathbf{x}_{m,n}^T (\mathbf{C}_{m,n}^{-1} - \mathbf{A}_{m,n}) \mathbf{t}_{m,n} + \frac{1}{2} \mathbf{x}_{m,n}^T \mathbf{A}_{m,n} \mathbf{x}_{m,n} \right]. \quad (5.5)$$

To calculate the WE $H_{\phi_{m,n}}^w(f_{\mathbf{C}_{m,n}}^{\text{No}}) = \int_{\mathbb{R}^{n-m+1}} I_{\phi_n}^w(\mathbf{x}_{m,n}) f_{\mathbf{C}_{m,n}}^{\text{No}}(\mathbf{x}_{m,n}) d\mathbf{x}_{m,n}$, we employ Gaussian integration formulas:

$$\begin{aligned}
H_{\phi_{m,n}}^w(f_{m,n}) &= \int_{\mathbb{R}^{n-m+1}} \frac{\log \left[(2\pi)^{n-m+1} \det \mathbf{C}_{m,n} \right] + \mathbf{x}_{m,n}^T \mathbf{C}_{m,n}^{-1} \mathbf{x}_{m,n} \log e}{2 \left[(2\pi)^{n-m+1} \det \mathbf{C}_{m,n} \right]^{1/2}} \\
&\quad \times \exp \left[\frac{1}{2} \mathbf{t}_{m,n}^T (\mathbf{C}_{m,n}^{-1} - \mathbf{A}_{m,n}) \mathbf{t}_{m,n} \right. \\
&\quad \left. - \frac{1}{2} (\mathbf{x}_{m,n}^T - \mathbf{t}_{m,n}^T) (\mathbf{C}_{m,n}^{-1} - \mathbf{A}_{m,n}) (\mathbf{x}_{m,n} - \mathbf{t}_{m,n}) \right] d\mathbf{x}_{m,n} \\
&= \frac{\exp \left[\frac{1}{2} \mathbf{t}_{m,n}^T (\mathbf{C}_{m,n}^{-1} - \mathbf{A}_{m,n}) \mathbf{t}_{m,n} \right]}{2 \left[\det (\mathbf{I}_{m,n} - \mathbf{C}_{m,n} \mathbf{A}_{m,n}) \right]^{1/2}} \left\{ H(f_{m,n}) + \text{tr} \left(\mathbf{I}_{m,n} - \mathbf{A}_{m,n} \mathbf{C}_{m,n} \right)^{-1} \log e \right\}.
\end{aligned} \tag{5.6}$$

In the case $\mathbf{A} = 0$ we obtain

$$\phi_{m,n}(\mathbf{x}_{m,n}) = \exp \left(\mathbf{x}_{m,n}^T \mathbf{C}_{m,n}^{-1} \mathbf{t}_{m,n} \right), \tag{5.7}$$

$$\begin{aligned}
I_{\phi_{m,n}}^w(\mathbf{x}_{m,n}, f_{m,n}) &= \frac{1}{2} \left\{ H(f_{m,n}) - (n - m + 1) \log e \right. \\
&\quad \left. + \mathbf{x}_{m,n}^T \mathbf{C}_{m,n}^{-1} \mathbf{x}_{m,n} \log e \right\} \exp \left(\mathbf{x}_{m,n}^T \mathbf{C}_{m,n}^{-1} \mathbf{t}_{m,n} \right)
\end{aligned} \tag{5.8}$$

and

$$H_{\phi_{m,n}}^w(f_{m,n}) = H(f_{m,n}) \exp \left(\frac{1}{2} \mathbf{t}_{m,n}^T \mathbf{C}_{m,n}^{-1} \mathbf{t}_{m,n} \right). \tag{5.9}$$

We arrive at a transparent conclusion. For a WF $\phi_n(\mathbf{x}_{0,n-1}) = \exp \left(\mathbf{x}_{0,n-1}^T \mathbf{C}_n^{-1} \mathbf{t}_{0,n-1} \right)$ (assuming $\mathbf{t} = (t_n : n \in \mathbb{Z})$ fixed), and given a sequence $(a(n), n \in \mathbb{Z}_+)$, the quantity

$$\begin{aligned}
J_n(\mathbf{x}_{0,n-1}) &:= \left[\frac{2I_{\phi_n}^w(\mathbf{x}_{0,n-1}, f_n)}{\phi_n(\mathbf{x}_{0,n-1})} + n \log e - \mathbf{x}_{0,n-1}^T \mathbf{C}_n^{-1} \mathbf{x}_{0,n-1} \right] \\
&\text{is a constant equal to } H(f_n) \text{ and hence } \frac{H(f_n)}{a(n)} \rightarrow \alpha \text{ iff } \frac{J_n}{a(n)} \rightarrow \alpha,
\end{aligned}$$

and

$$\frac{H(f_n)}{a(n)} \rightarrow \alpha \text{ iff } \frac{H_{\phi_n}^w(f_n)}{a(n)} \exp \left(-\frac{1}{2} \mathbf{t}_{m,n}^T \mathbf{C}_{m,n}^{-1} \mathbf{t}_{m,n} \right) \rightarrow \alpha.$$

On the other hand, for $\mathbf{t} = \mathbf{0}$, the WF is simplified to

$$\phi_{m,n}(\mathbf{x}_{m,n}) = \exp \left(\frac{1}{2} \mathbf{x}_{m,n}^T \mathbf{A}_{m,n} \mathbf{x}_{m,n} \right) \tag{5.10}$$

whereas the WI and WE, respectively, to

$$\begin{aligned}
I_{\phi_{m,n}}^w(\mathbf{x}_{m,n}, f_{m,n}) &= \frac{1}{2} \left\{ H(f_{m,n}) - (n - m + 1) \log e \right. \\
&\quad \left. + \mathbf{x}_{m,n}^T \mathbf{C}_{m,n}^{-1} \mathbf{x}_{m,n} \log e \right\} \exp \left(\frac{1}{2} \mathbf{x}_{m,n}^T \mathbf{A}_{m,n} \mathbf{x}_{m,n} \right).
\end{aligned} \tag{5.11}$$

Furthermore,

$$H_{\phi_{m,n}}^w(f_{m,n}) = \frac{H(f_{m,n}) + \text{tr} \left[\left(\mathbf{I}_{m,n} - \mathbf{A}_{m,n} \mathbf{C}_{m,n} \right)^{-1} \right] \log e}{2 \left[\det \left(\mathbf{I}_{m,n} - \mathbf{C}_{m,n} \mathbf{A}_{m,n} \right) \right]^{1/2}}. \quad (5.12)$$

This implies that, for $\phi_n(\mathbf{x}_{0,n-1}) = \exp \left(\frac{1}{2} \mathbf{x}_{0,n-1}^T \mathbf{A}_n \mathbf{x}_{0,n-1} \right)$, the map $\mathbf{x}_{0,n-1} \mapsto K_n(\mathbf{x}_{0,n-1})$ yields a constant equal to $H(f_{\mathbf{C}_n}^{\text{No}})$. Here K_n has an expression analogous to J_n :

$$K_n := \left[\frac{2I_{\phi_n}^w(\mathbf{x}_{0,n-1}, f_n)}{\phi_n(\mathbf{x}_{0,n-1})} + n \log e - \mathbf{x}_{0,n-1}^T \mathbf{C}_n^{-1} \mathbf{x}_{0,n-1} \right],$$

and hence $\frac{H(f_n)}{a(n)} \rightarrow \alpha$ iff $\frac{K_n}{a(n)} \rightarrow \alpha$.

Also,

$$\frac{H(f_n)}{a(n)} \rightarrow \alpha \text{ iff } \frac{1}{a(n)} \left\{ 2H_{\phi_n}^w(f_n) \left[\det \left(\mathbf{I}_n - \mathbf{C}_n \mathbf{A}_n \right) \right]^{1/2} - \text{tr} \left[\left(\mathbf{I}_n - \mathbf{A}_n \mathbf{C}_n \right)^{-1} \right] \log e \right\} \rightarrow \alpha.$$

Similar manipulations can be performed in the general case.

Acknowledgement

YS thanks the Math. Department, Penn State University, for hospitality and support. IS thanks the Math. Department, University of Denver, for support and hospitality.

References

- [1] P. Algoet, T. Cover. A sandwich proof of the Shannon–McMillan–Breiman theorem. *Ann. Probab.*, **16**, No. 2 (1988), 899–909.
- [2] M. Arato. *Linear Stochastic Systems with Constant Coefficients. A Statistical Approach*. Berlin et al.: Springer-Verlag, 1982.
- [3] A.R. Barron. The strong ergodic theorem for densities: generalized Shannon–McMillan–Breiman theorem. *Ann. Prob.*, **13** (1985), 1292–1303.
- [4] T. Cover, J. Thomas. *Elements of Information Theory*. New York: Wiley, 2006.
- [5] K. Deimling. *Nonlinear Functional Analysis*. Berlin et al.: Springer, 1985; reprinted Mineola, NY: Dover, 2010.
- [6] H.-O. Georgii. *Gibbs Measures and Phase Transitions*. Berlin/New York: De Gruyter, 2011.

- [7] I. A. Ibragimov, Y. A. Rozanov. *Gaussian Random Processes*. New York et al.: Springer-Verlag, 1978.
- [8] S. Kalikow, R. McCutcheon. *An Outline of Ergodic Theory*. Cambridge; CUP, 2010.
- [9] M.G. Krein, M. A. Rutman. Linear operators leaving invariant a cone in a Banach space". *Amer. Math. Soc. Transl.*, 1950, No 26.
- [10] M. Kelbert and Y. Suhov. *Information Theory and Coding by Example*. Cambridge: CUP, 2013.
- [11] D. Ruelle. Statistical mechanics of a one-dimensional lattice gas. *Comm. Math. Phys.*, **9**:4 (1968), 267-278.
- [12] D. Ruelle. *Statistical mechanics. Rigorous results*. London: Imperial College Press; Singapoer: World Scientific, 1999.
- [13] D. Ruelle. *Thermodynamic formalism*. Cambridge: CUP, 2004
- [14] O. Sarig. *Lecture Notes on Ergodic Theory*. www.wisdom.weizmann.ac.il/~sarigo, 2009.
- [15] Y. Suhov, I. Stuhl, On principles of large deviation and selected data compression. arXiv:1604.06971
- [16] Y. Suhov, I. Stuhl, M. Kelbert. Weight functions and log-optimal investment portfolios. 1505.01437
- [17] Y. Suhov, I. Stuhl, S. Yasaei Sekeh, M. Kelbert. Basic inequalities for weighted entropies. *Aequationes Math.*, **90** (2016), 817-848.
- [18] Y. Suhov, I. Stuhl, S. Yasaei Sekeh. Weighted Gaussian entropy and determinant inequalities. arXiv:1505.01753v1.
- [19] Yu. Suhov. The matrix method for continuous systems of statistical mechanics. *Trudy Mosc. Matem. Ob.* (Russian), **24** (1971), 175–200; English translation: *Trans. Moscow Math. Soc.*, **24** (1971), 185-212.
- [20] Y. Suhov. Random point processes and DLR equations. *Commun. Math. Phys.* **50** (1976), 113–131.
- [21] P. Walters. *An introduction to ergodic theory*. New York: Springer-Verlag, 1982.